



A convex approach to superresolution and regularization of lines in images

Kevin Polisano, Laurent Condat, Marianne Clausel, Valérie Perrier

► To cite this version:

Kevin Polisano, Laurent Condat, Marianne Clausel, Valérie Perrier. A convex approach to super-resolution and regularization of lines in images. SIAM Journal on Imaging Sciences, 2019, 12 (1), pp.211-258. 10.1137/18M118116X . hal-01599010v4

HAL Id: hal-01599010

<https://hal.science/hal-01599010v4>

Submitted on 19 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A CONVEX APPROACH TO SUPER-RESOLUTION AND REGULARIZATION OF LINES IN IMAGES

KÉVIN POLISANO*, LAURENT CONDAT†, MARIANNE CLAUSEL*, AND VALÉRIE
PERRIER*

Abstract. We present a new convex formulation for the problem of recovering lines in degraded images. Following the recent paradigm of super-resolution, we formulate a dedicated atomic norm penalty and we solve this optimization problem by means of a primal-dual algorithm. This parsimonious model enables the reconstruction of lines from lowpass measurements, even in presence of a large amount of noise or blur. Furthermore, a Prony method performed on rows and columns of the restored image, provides a spectral estimation of the line parameters, with subpixel accuracy.

Key words. Super-resolution, sparse recovery, convex optimization, line detection, splitting method, spectral estimation

AMS subject classifications. 94A08, 94A12, 94A20, 47N10, 90C22, 90C25

1. Introduction. Many restoration or reconstruction imaging problems are ill-posed and must be regularized. So, they can be formulated as convex optimization problems formed by the combination of a data fidelity term with a norm-based regularizer. Typically, given the data $\mathbf{y} = \mathbf{A}\mathbf{x}^\sharp + \boldsymbol{\epsilon}$, for some unknown image \mathbf{x}^\sharp to estimate, some known observation operator \mathbf{A} and some noise $\boldsymbol{\epsilon}$, one aims at solving a problem like:

$$\text{Find } \tilde{\mathbf{x}} \in \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \lambda R(\mathbf{x}),$$

where λ controls the tradeoff between data fidelity and regularization and R is a convex regularization functional. R can be chosen to promote some kind of smoothness. The classical Tikhonov regularizer $R(\mathbf{x}) = \|\nabla \mathbf{x}\|_2^2$ generally makes the problem easy to solve, but yields over-smoothing of the textures and edges in the recovered image $\tilde{\mathbf{x}}$. A popular and better regularizer is the total variation $R(\mathbf{x}) = \|\nabla \mathbf{x}\|_1$, see e.g [16, 26]; it yields images with sharp edges, but the textures are still over-smoothed, there are staircasing effects and the pixel values tend to be clustered in piecewise constant areas. To overcome these drawbacks, one can penalize higher order derivatives [50, 9] or make use of nonlocal penalties [63, 24, 21]. Another approach, which is at the heart of the recent paradigm of sparse recovery [77, 41] and compressed sensing [35, 81], is to choose R to favor some notion of low complexity. Indeed, many phenomena, when observed by instruments, yield data living in high dimensional spaces, but inherently governed by a small number of degrees of freedom. One early choice was to set R as the ℓ_1 norm of wavelet coefficients of the image. But the signals encountered in applications like radar, array processing, communication, seismology, or remote sensing, are usually specified by parameters in a *continuous* domain, from which they depend nonlinearly. So, modern sampling theory has widened its scope to a broader class of signals, with so-called finite rate of innovation, i.e. ruled by parsimonious models [52, 37, 6, 85]. This encompasses reconstruction of pulses from lowpass measurements [28] and spectral estimation, which is the reconstruction of sinusoids from point sam-

*Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France.

E-mail: kevin.polisano@univ-grenoble-alpes.fr.

†Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France.

E-mail: laurent.condat@gipsa-lab.grenoble-inp.fr.

ples [78, 79], with many applications [18, 47, 84, 10, 80, 51, 69, 76, 27]. The knowledge of the kind of elements we want to promote in the image makes it possible to estimate them from coarse-scale measurements, even with infinite precision if there is no noise. Methods achieving this goal are qualified as *super-resolution* methods, because they uncover fine scale information, which was lost in the data, beyond the Rayleigh or Nyquist resolution limit of the acquisition system [43, 13]. However, in this context, maximum likelihood estimation amounts to structured low rank approximation, which forms nonconvex and very difficult, even NP-hard in general, problems [55, 28]. An elegant and unifying formulation, which yields convex problems, is based on the *atomic norm* [4, 25]. We place ourselves in this general framework of atomic norm minimization: the sought-after image \mathbf{x}^\sharp is supposed to be a sparse positive combination of the elements of an infinite dictionary \mathcal{A} , indexed by continuously varying parameters. Then, one can choose R as the atomic norm $\|\mathbf{x}\|_{\mathcal{A}}$ of the image \mathbf{x} , which can be viewed as the ℓ_1 norm of the coefficients, when the image is expressed in terms of the unit-norm elements of \mathcal{A} , called *atoms*:

$$(1) \quad \|\mathbf{x}\|_{\mathcal{A}} = \inf \{t > 0 : \mathbf{x} \in t \operatorname{conv}(\mathcal{A})\},$$

where $\operatorname{conv}(\mathcal{A})$ is the convex hull of the atoms. In this paper, we consider the setting, which is new to our knowledge, where the atoms are *lines*. Expressed in the Fourier domain, these atoms can be characterized with respect to their rows and columns and the problem can be reduced to a dictionary of 1-D complex exponential samples, indexed by their frequency and phase, whose atomic norms can be computed via semidefinite programming [86]. This formulation makes it possible to derive a convex optimization problem under constraints, solved by mean of a primal-dual splitting algorithm [23]. Then, applying a Prony-like method [71] to the solution of the algorithm allows us to extract the parameters of the lines. This approach estimates the lines with high accuracy, whereas the Hough [45, 46, 56] and the Radon [70, 31, 57] transforms fail, due to their discrete nature. Our motivation stems from the frequent presence in biomedical images, e.g. in microscopy, of elongated structures like filaments, neurons, veins, which are deteriorated when reconstructed with classical penalties.

Some related works are dedicated to the recovery of curve-like singularities, by variational methods [1, 88], Riesz-based models [53, 54] or so-called finite rate of innovation methods [52, 75, 20, 61, 42, 59, 61]. The originality of our method is that we reduce the minimization over an infinite dictionary of lines to a semidefinite programming problem, taking advantage of the line structure in both directions of the grid. Although there are works in the same vein to recover 2-D point sources, or equivalently to estimate the parameters of 2-D exponentials [67, 64, 29, 36], applying similar principles to the estimation of lines is not straightforward and is new, to our knowledge.

The paper is organized as follows. The model is described in [section 2](#), the framework of atomic norm minimization underlying the super-resolution principle is introduced in [section 3](#), the algorithms we derive are presented in [section 4](#). Then a Prony-like method is developed in [section 5](#), as a way to perform spectral estimation of the line parameters. [Section 6](#) gives an overview of related works. Finally some experimental results are shown in [section 7](#). Part of this work has been published in a conference paper [66]. In the present paper, we added mathematical developments, another algorithm, a new estimation procedure for the line parameters, an extension to the whole range of line angles without restriction any more, an application to inpainting problems and several numerical experiments.

Notations. Vectors, e.g. $\mathbf{z} = (z_0, \dots, z_{N-1})$ and matrices have component indexes starting at zero. The Hilbert space of complex matrices of size $M \times N$ is denoted by $\mathcal{M}_{M,N}$. The entry in the $(k_1 + 1)$ -th row and $(k_2 + 1)$ -th column of a matrix \mathbf{M} is referred to $\mathbf{M}[k_1, k_2]$ or \mathbf{M}_{k_1, k_2} . Due to required multiple subscripts, we often adopt the MATLABTM notation $\mathbf{M}[k_1, :]$ (resp. $\mathbf{M}[:, k_2]$) for referring to the $(k_1 + 1)$ -th row (resp. $(k_2 + 1)$ -th column) of the matrix \mathbf{M} . To extract a submatrix, we also use the notation $\mathbf{M}[p_1 : q_1, p_2 : q_2]$. The nuclear norm of the matrix \mathbf{M} , denoted by $\|\mathbf{M}\|_* = \sum_i \sigma_i(\mathbf{M})$, is the sum of its singular values $\sigma_i(\mathbf{M})$. The operation $\mathbf{M} \otimes \mathbf{N}$ denotes the Kronecker product between a $m \times n$ matrix \mathbf{M} and a $p \times q$ matrix \mathbf{N} , whose result is the $mp \times nq$ matrix

$$(2) \quad \mathbf{M} \otimes \mathbf{N} = \begin{pmatrix} \mathbf{M}_{0,0}\mathbf{N} & \cdots & \mathbf{M}_{0,n-1}\mathbf{N} \\ \vdots & \ddots & \vdots \\ \mathbf{M}_{m-1,0}\mathbf{N} & \cdots & \mathbf{M}_{m-1,n-1}\mathbf{N} \end{pmatrix}.$$

The ground truths for distributions or matrices to recover are denoted by a sharp \cdot^\sharp , while the estimated variables are denoted by a tilde $\tilde{\cdot}$. Continuous (resp. discrete) 1-D Fourier transform of functions or distributions (resp. vectors) is denoted by a hat $\hat{\cdot}$. For 2-D functions, distributions, or matrices, the notation \hat{f} always refers to the horizontal Fourier transform of f , while the 2-D Fourier transform is denoted by $\mathcal{F}f$. When it is necessary to distinguish them, the horizontal and vertical Fourier transform are denoted by $\mathcal{F}_1 f$ and $\mathcal{F}_2 f$, respectively. The Radon transform of a function f is denoted by $\mathcal{R}f$. The projection operator onto a set C is denoted by P_C and the proximity operator of a closed proper convex function $f : \mathbb{R}^N \rightarrow \mathbb{R} \cup \{+\infty\}$ (see [8, 3, 22, 62]) is defined by

$$(3) \quad \text{prox}_f(\mathbf{x}) = \arg \min_{\mathbf{y}} \left(f(\mathbf{y}) + \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 \right).$$

2. An image model of blurred lines. Our aim is to restore lines from an observed image $\mathbf{y} = \mathbf{b}^\sharp + \epsilon$, which is made of a blurred image \mathbf{b}^\sharp containing lines and corrupted by some noise ϵ and then to estimate the parameters—angle, offset, amplitude—of the lines. In this section, we formulate what we precisely mean by an image containing lines. In short, \mathbf{b}^\sharp is a sum of perfect lines which have been blurred and then sampled. Both processes are detailed in the following.

2.1. The ideal continuous model and the objectives. We place ourselves in the quotient space $\mathbb{P} = \mathbb{R}/(W\mathbb{Z}) \times \mathbb{R}$, corresponding to the 2-D plane with horizontal W -periodicity, for some integer $W \geq 1$. To simplify the notations, we suppose that W is odd and we set $M = (W - 1)/2$.

A perfect line, with angle $\theta \in (-\pi/2, \pi/2]$ with respect to the vertical axis, amplitude $\alpha > 0$ and offset $\eta \in \mathbb{R}$ from the origin on the horizontal axis, can be defined as a tempered distribution, which maps a function ψ in the Schwartz class $\mathcal{S}(\mathbb{R}^2)$ to its integral along the geometric line $\mathcal{L} = \{(t_1, t_2) \in \mathbb{P} : (t_1 - \eta) \cos \theta + t_2 \sin \theta = 0\}$; that is, to $\alpha \int_{\mathcal{L}} \psi(t_1, t_2) dt_1 dt_2$. Thus, by a slight abuse of notation, we can write the perfect line as:

$$(t_1, t_2) \in \mathbb{P} \mapsto \alpha \delta((t_1 - \eta) \cos \theta + t_2 \sin \theta),$$

where δ is the (1-D) Dirac distribution. For more details about multi-dimensional distributions over curves or surfaces, we refer the interested reader to [44, 60].

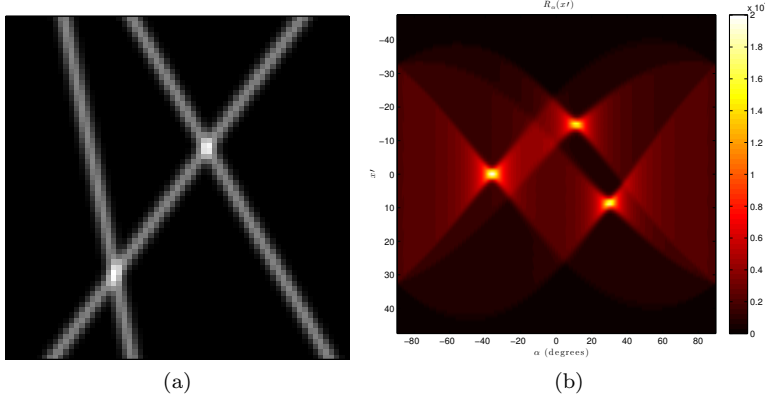


FIG. 1. (a) The image \mathbf{b}^\sharp of three blurred lines and (b) the Radon transform of \mathbf{b}^\sharp .

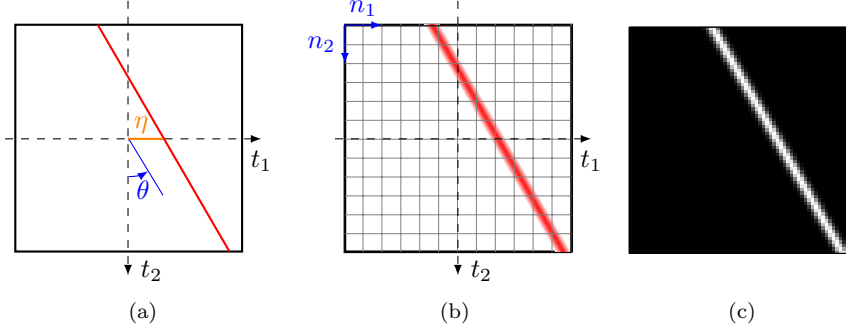


FIG. 2. (a) Parameters (θ, η) characterizing the position of a line in the 2-D plane, (b) the matrix convention we use to display the image obtained by sampling with unit step the blurred line $\mathbf{b}^\sharp = x^\sharp * \phi$ and (c) the resulting discrete image $\mathbf{b}^\sharp[n_1, n_2] = (x^\sharp * \phi)(n_1, n_2)$.

We define the distribution x^\sharp , which is a sum of K different such perfect lines, for some integer $K \geq 1$, as

$$(4) \quad x^\sharp : (t_1, t_2) \in \mathbb{P} \mapsto \sum_{k=1}^K \alpha_k \delta((t_1 - \eta_k) \cos \theta_k + t_2 \sin \theta_k) .$$

Figure 2a illustrates the line parameters and Figure 2b the convention we use for representing images.

HYPOTHESIS 1. *At this time, we suppose that the lines are rather vertical; that is, for every $k = 1, \dots, K$, $\theta_k \in (-\pi/4, \pi/4]$.*

This hypothesis is made, because the rows and columns of the image will be processed differently. We will proceed in the Fourier domain, by applying the discrete 1-D Fourier transform on every row of the image; we insist on the fact that we do not consider the usual 2-D Fourier transform. This setting may appear restrictive, but we show in subsection 4.3 that, in fact, we are able to deal with the general case.

2.2. A blur model for an exact sampling process. The observed image \mathbf{b}^\sharp of size $W \times H$ in Figure 2c is obtained by convolution of the distribution x^\sharp with a

blur function ϕ , followed by sampling with unit step:

$$(5) \quad \mathbf{b}^\sharp[n_1, n_2] = (x^\sharp * \phi)(n_1, n_2), \quad \forall n_1 = 0, \dots, W-1, \quad n_2 = 0, \dots, H-1.$$

The point spread function (PSF) ϕ is assumed to be separable; that is, it can be written $\phi(t_1, t_2) = \varphi_1(t_1)\varphi_2(t_2)$. So, the function $b^\sharp = x^\sharp * \phi$ can be obtained by a first horizontal convolution with φ_1 and then a second vertical convolution with φ_2 . We can show (see [Appendix A](#)) that we get the function

$$(6) \quad b^\sharp = x^\sharp * \phi : (t_1, t_2) \in \mathbb{P} \mapsto \sum_{k=1}^K \alpha_k \psi_k((t_1 - \eta_k) \cos \theta_k + t_2 \sin \theta_k),$$

where

$$(7) \quad \psi_k = \left(\frac{1}{\cos \theta_k} \varphi_1\left(\frac{\cdot}{\cos \theta_k}\right) \right) * \left(\frac{1}{\sin \theta_k} \varphi_2\left(\frac{\cdot}{\sin \theta_k}\right) \right),$$

if $\theta_k \neq 0$ and $\psi_k = \varphi_1$ else.

REMARK 1. We can notice that (6) can also be interpreted as follows: every line has undergone a 1-D convolution with ψ_k in the direction transverse to it. We can also notice that if φ_1 and φ_2 are Gaussian functions and have same variance κ^2 , it follows from (7) that ψ_k has variance $\kappa^2(\cos^2 \theta + \sin^2 \theta) = \kappa^2$ as well.

Assumptions. We assume that φ_1 and φ_2 have the following properties:

- (i) $\varphi_1 \in L^1(0, W)$ is W -periodic, bounded, such that $\frac{1}{W} \int_0^W \varphi_1(t_1) dt_1 = 1$ and bandlimited; that is, its Fourier coefficients

$$c_m(\varphi_1) = \frac{1}{W} \int_0^W \varphi_1(t_1) e^{-j2\pi m t_1 / W} dt_1$$

are zero for every $m \in \mathbb{Z}$ with $|m| \geq (W+1)/2 = M+1$. Then, the discrete filter

$$(8) \quad (\mathbf{g}[n] = \varphi_1(n))_{n \in \mathbb{Z}}$$

has discrete Fourier coefficients

$$\hat{\mathbf{g}}[m] = \frac{1}{W} \sum_{n=0}^{W-1} \mathbf{g}[n] e^{-j \frac{2\pi m n}{W}} = c_m(\varphi_1).$$

- (ii) $\varphi_2 \in L^1(\mathbb{R})$ is such that $\int_{\mathbb{R}} \varphi_2(t_2) dt_2 = 1$ and bandlimited so that the discrete filter

$$(9) \quad (\mathbf{h}[n] = \varphi_2(n))_{n \in \mathbb{Z}}$$

has compact support of length $2S+1$, for some $S \in \mathbb{N}$; that is,

$$\mathbf{h}[n] = 0, \quad \text{if } |n| \geq S+1.$$

An example of three blurred lines is depicted in [Figure 1a](#). We insist on the fact that no discrete approximation is made during the sampling process leading from the continuous to the discrete formulation, due to the Assumptions (i), (ii) and [Hypothesis 1](#). These important guarantees, for the purpose of super-resolution, are detailed in [Appendix A](#).

REMARK 2. *These assumptions are convenient and reasonable. Indeed, in practice, we can always approximate a PSF by a bandlimited function. That said, we also propose in [Appendix A](#) a weaker version of Assumption (ii), denoted by (ii'), which relax the bandlimited restriction. Regarding the periodicity in Assumption (i), this hypothesis can be circumvented by considering an acquisition process yielding a twice larger image, in order to “periodize” the observed image.*

2.3. Toward an inverse problem in the Fourier domain. Let us further characterize the blurred image \mathbf{b}^\sharp in Fourier domain, i.e. its 1-D discrete Fourier transform (DFT) on its first component (in the horizontal direction) denoted by $\widehat{\mathbf{b}}^\sharp$. To achieve this, let us also denote by $\widehat{\mathbf{x}}^\sharp$ the discrete image obtained by sampling $\widehat{x}^\sharp = \mathcal{F}_1 x^\sharp$, which is the horizontal Fourier transform, in the sense of distributions, of the ideal model x^\sharp (4) made up of 1-D Dirac distributions. Therefore, $\widehat{\mathbf{x}}^\sharp$ is composed of a sampled sum of exponentials:

$$(10) \quad \widehat{\mathbf{x}}^\sharp[m, n_2] = \sum_{k=1}^K \frac{\alpha_k}{\cos \theta_k} e^{j2\pi(n_2 \tan \theta_k - \eta_k) \frac{m}{W}},$$

$$m = -M, \dots, M, \quad n_2 = -S, \dots, H-1+S.$$

REMARK 3. $\widehat{\mathbf{x}}^\sharp$ is such that $\widehat{\mathbf{x}}^\sharp[-m, n_2] = \widehat{\mathbf{x}}^\sharp[m, n_2]^*$, where \cdot^* denotes the complex conjugation, so we can only deal with the right part $\widehat{\mathbf{x}}^\sharp[0 : M, :]$ and we can note that the row corresponding to $m = 0$ is real and equal to $\sum_{k=1}^K \frac{\alpha_k}{\cos \theta_k}$. We consider in the following the Fourier image $\widehat{\mathbf{x}}^\sharp[m, n_2]$ of size $(M+1) \times H_S$, with $H_S = H + 2S$, due to the addition of S pixels beyond the boundaries, allowing the convolution with filter \mathbf{h} determined hereafter.

Our goal will be to reconstruct $\widehat{\mathbf{x}}^\sharp$ from $\widehat{\mathbf{b}}^\sharp$, that is from its observations through a known degradation operator \mathbf{A} , which we characterize in the following proposition:

PROPOSITION 1. *Let \mathbf{A} be the operator which for all $m = 0, \dots, M$ multiplies each row vector $\widehat{\mathbf{x}}^\sharp[m, :]$ of $\widehat{\mathbf{x}}^\sharp$ by the corresponding Fourier coefficient $\widehat{\mathbf{g}}[m]$ and convolves it with the filter $\mathbf{h} = (h_{-S}, \dots, h_0, \dots, h_S)$. Then, we have the relation*

$$(11) \quad \mathbf{A} \widehat{\mathbf{x}}^\sharp = \widehat{\mathbf{b}}^\sharp,$$

which can be alternatively rewrite on each row $m = 0, \dots, M$ as:

$$(12) \quad \widehat{\mathbf{g}}[m](\widehat{\mathbf{x}}^\sharp[m, :] * \mathbf{h}) = \widehat{\mathbf{b}}^\sharp[m, :].$$

Proof. See in [Appendix B](#). □

REMARK 4. *From a matrix point of view, the operator \mathbf{A} corresponds to a left and right matrix multiplication with the matrices $\widehat{\mathbf{G}}$ of size $(M+1) \times (M+1)$ and the transpose of $\check{\mathbf{H}}$ of size $H \times H_S$ defined by:*

$$(13) \quad \widehat{\mathbf{G}} = \text{diag}(\widehat{g}_0, \dots, \widehat{g}_M), \quad \check{\mathbf{H}} = \begin{pmatrix} h_{-S} & \cdots & h_S & 0 & 0 & \cdots & 0 \\ 0 & h_{-S} & \cdots & h_S & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & \ddots & \ddots & \cdots & \vdots \\ 0 & \cdots & 0 & h_{-S} & \cdots & h_S & 0 \\ 0 & \cdots & 0 & 0 & h_{-S} & \cdots & h_S \end{pmatrix};$$

that is,

$$(14) \quad \mathbf{A} \widehat{\mathbf{x}}^\sharp = \widehat{\mathbf{G}} \widehat{\mathbf{x}}^\sharp \check{\mathbf{H}}^\top.$$

Finally, the image \mathbf{b}^\sharp of the blurred lines is corrupted by noise, so that we observe the degraded image

$$(15) \quad \mathbf{y} = \mathbf{b}^\sharp + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \zeta),$$

with ζ the noise level. Thus, the problem consists from (15) and (11) in recovering $\hat{\mathbf{x}}$ from $\hat{\mathbf{y}}$, which takes the form of a ill-posed linear inverse problem. Then, we will need to express a convex optimization problem, under constraints exploiting the sparse structure of the signal we are looking for, namely that it is a combination of lines. The super-resolution process consists of recovering the high frequency content (lost because of the blur operator) from the degraded image \mathbf{y} , which can be viewed as a spectral extrapolation process. Then, we aim at recovering the parameters $(\theta_k, \eta_k, \alpha_k)$ of these lines, as a post-process, after the reconstruction. This procedure will be decomposed as:

1. First, solve a convex optimization problem of the form:

$$(16) \quad \text{Minimize } \|\hat{\mathbf{y}} - \mathbf{A}\hat{\mathbf{x}}\|, \text{ under the constraint that } \hat{\mathbf{x}} \text{ is made of lines ;}$$

that is, to go to the bottom line of the diagram in Figure 15 from \mathbf{y} to $\hat{\mathbf{x}}^\sharp$.

2. Second, perform a Prony-like method on $\hat{\mathbf{x}}^\sharp$ in order to estimate the K parameters $(\theta_k, \eta_k, \alpha_k)$.

These two steps are summarized in Figure 3. Note that this work also covers the case where a mask is applied; that is, it can encompass inpainting problems. In the next section, we present the framework of atomic norm, from which the optimization problem will be derived.

3. Super-resolution detection of lines.

3.1. Atomic norm and semidefinite characterizations. Consider a complex signal $\mathbf{z} \in \mathbb{C}^N$ represented as a K -sparse mixture of atoms from the set

$$\mathcal{A} = \{\mathbf{a}(\omega) \in \mathbb{C}^N : \omega \in \Omega\},$$

that is,

$$\mathbf{z} = \sum_{k=1}^K c_k \mathbf{a}(\omega_k), \quad c_k \geq 0, \quad \omega_k \in \Omega.$$

We consider atoms $\mathbf{a}(\omega) \in \mathbb{C}^N$ that are continuously indexed in the dictionary \mathcal{A} by the parameter ω in a compact set Ω . The atomic norm, first introduced in [19], is defined as

$$\|\mathbf{z}\|_{\mathcal{A}} = \inf \{t > 0 : \mathbf{z} \in t \text{conv}(\mathcal{A})\},$$

where $\text{conv}(\mathcal{A})$ denotes the convex hull of a general atomic set \mathcal{A} , enforcing sparsity. Chandrasekaran et al. [19] argue that the atomic norm is the best convex heuristic for underdetermined, structured linear inverse problems, which generalizes the ℓ_1 norm for sparse recovery and the nuclear norm for low-rank matrix completion.

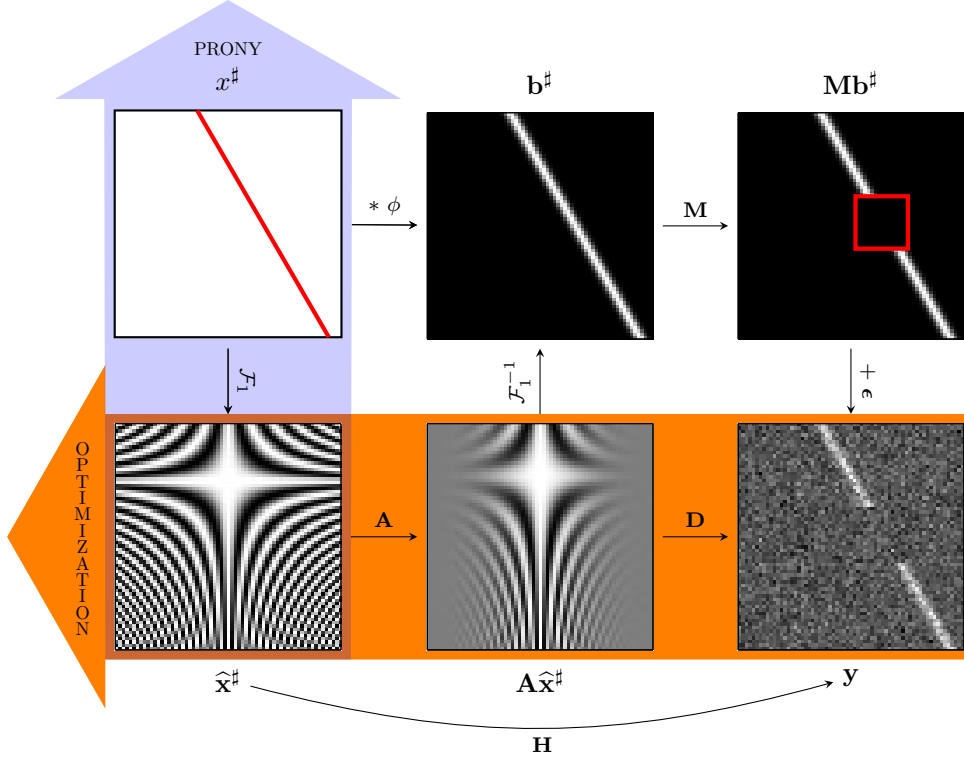


FIG. 3. The two steps of the procedure : a convex optimization formulation for the reconstruction of the lines (in orange) and a Prony-like method for the estimation of their parameters (in purple).

In our problem (4) the atoms are somehow *lines*, so one can consider in the Fourier domain the dictionary \mathcal{A}_{2D} indexed by the angle and the offset; that is, composed of the 2-D exponential atoms of size $W \times H_S$:

$$\mathbf{a}_{2D}(\theta_k, \eta_k) = \frac{1}{\cos \theta_k} e^{j2\pi(n_2 \tan \theta_k - \eta_k)m/W},$$

$$m = -M, \dots, M, \quad n_2 = -S, \dots, H - 1 + S,$$

as illustrated in Figure 4. The problem is that there is no closed-form expression for the atomic norm in these 2-D dictionaries, to our knowledge. However, in the case of 1-D complex exponentials, there is a way to compute the atomic norm via semidefinite programming. So, we reformulate the problem using the simplified 1-D case. From now on, we consider the dictionaries

$$(17) \quad \mathcal{A} = \left\{ \mathbf{a}(f, \phi) \in \mathbb{C}^{|I|}, \quad f \in [0, 1], \quad \phi \in [0, 2\pi) \right\},$$

$$(18) \quad \mathcal{A}_0 = \left\{ \mathbf{a}(f) \in \mathbb{C}^{|I|}, \quad f \in [0, 1] \right\},$$

in which the atoms are the vectors of components $[\mathbf{a}(f, \phi)]_i = e^{j(2\pi f i + \phi)}$, $i \in I$ and

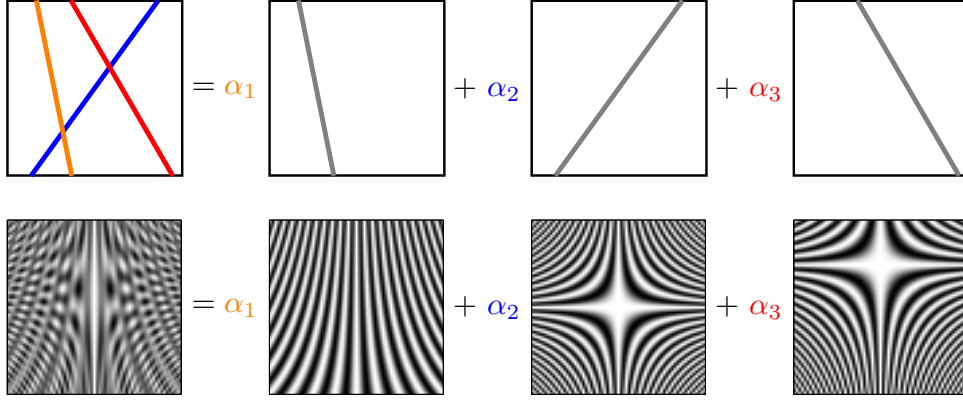


FIG. 4. Illustration with a signal made of a weighted combination of three lines (in gray). In the Fourier domain, we have through the sampling process the same kind of combination, but with 2-D exponential atoms $\mathbf{a}_{2D}(\theta_k, \eta_k)$. In both cases, the weights α_i are the amplitudes of the lines.

$[\mathbf{a}(f)]_i = [\mathbf{a}(f, 0)]_i = e^{j2\pi f i}$, $i \in I$. The atomic norm writes:

$$(19) \quad \|\mathbf{z}\|_{\mathcal{A}} = \inf_{\substack{c'_k \geq 0 \\ f'_k \in [0,1) \\ \phi'_k \in [0,2\pi)}} \left\{ \sum_k c'_k : \mathbf{z} = \sum_k c'_k \mathbf{a}(f'_k, \phi'_k) \right\}.$$

THEOREM 2 (Caratheodory). A vector $\mathbf{z} = (z_{N-1}^*, \dots, z_1^*, z_0, z_1, \dots, z_{N-1})$ of length $2N - 1$, with $z_0 \in \mathbb{R}$, is a positive combination of $K \leq N$ atoms $\mathbf{a}(f_k)$ if and only if $\mathbf{T}_N(\mathbf{z}_+) \succcurlyeq 0$ and is of rank K , where $\mathbf{z}_+ = (z_0, \dots, z_{N-1})$ is of length N , $\mathbf{T}_N : \mathbb{C}^N \rightarrow \mathcal{T}_N$ is the Toeplitz operator

$$(20) \quad \mathbf{T}_N : \mathbf{z}_+ = (z_0, \dots, z_{N-1}) \mapsto \begin{pmatrix} z_0 & z_1^* & \cdots & z_{N-1}^* \\ z_1 & z_0 & \cdots & z_{N-2}^* \\ \vdots & \vdots & \ddots & \vdots \\ z_{N-1} & z_{N-2} & \cdots & z_0 \end{pmatrix},$$

with \mathcal{T}_N the Hilbert subspace of \mathcal{M}_N composed of Hermitian Toeplitz matrices and $\succcurlyeq 0$ denotes positive semidefiniteness. Moreover, this decomposition is unique, if $K < N$.

Proof. See references [14, 15, 83, 65]. \square

We are now in position to characterize the atomic norm:

PROPOSITION 3. The atomic norm $\|\mathbf{z}\|_{\mathcal{A}}$ can be characterized by the following semidefinite program:

$$(21) \quad \|\mathbf{z}\|_{\mathcal{A}} = \min_{\mathbf{q} \in \mathbb{C}^N, q_0 \geq 0} \left\{ q_0 : \mathbf{T}'_N(\mathbf{z}, \mathbf{q}) = \begin{pmatrix} \mathbf{T}_N(\mathbf{q}) & \mathbf{z} \\ \mathbf{z}^* & q_0 \end{pmatrix} \succcurlyeq 0 \right\},$$

where q_0 is the first component of vector $\mathbf{q} = (q_0, \dots, q_{N-1}) \in \mathbb{R}^+ \times \mathbb{C}^{N-1}$ and $\mathbf{z}^* = \bar{\mathbf{z}}^T$.

Proof. This result is an improvement of [82, Proposition II.1] and the proof is given in [Appendix C](#). \square

Since the matrix $\mathbf{T}'_N(\mathbf{z}, \mathbf{q})$ in (21) is Hermitian and positive semidefinite, its eigenvalues $(\lambda_i)_{0 \leq i \leq N}$ are positive reals. So, $q_0 = \frac{1}{N+1} \text{tr}(\mathbf{T}'_N(\mathbf{z}, \mathbf{q})) = \frac{1}{N+1} \sum_{i=0}^N \lambda_i$ is real and positive.

3.2. Properties of the model $\hat{\mathbf{x}}^\sharp$ with respect to the atomic norm. In Fourier domain, the discrete image $\hat{\mathbf{x}}^\sharp$ given by (10) can be viewed as a sum of atoms: regarding the columns $\mathbf{l}_{n_2}^\sharp$ of the matrix $\hat{\mathbf{x}}^\sharp$, with $I = \{0, \dots, H_S - 1\}$, we have:

$$(22) \quad \mathbf{l}_{n_2}^\sharp = \hat{\mathbf{x}}^\sharp[:, n_2] = \sum_{k=1}^K c_k \mathbf{a}(f_{n_2, k}) ,$$

and regarding the rows \mathbf{t}_m^\sharp , with $I = \{-M, \dots, M\}$, we have:

$$(23) \quad \mathbf{t}_m^\sharp = \hat{\mathbf{x}}^\sharp[m, :] = \sum_{k=1}^K c_k \mathbf{a}(f_{m, k}, \phi_{m, k})^\top ,$$

where

$$(24) \quad \begin{aligned} c_k &= \frac{\alpha_k}{\cos \theta_k}, \quad f_{n_2, k} = \frac{n_2 \tan \theta_k - \eta_k}{W}, \\ \phi_{m, k} &= -\frac{2\pi m \eta_k}{W}, \quad f_{m, k} = \frac{m \tan \theta_k}{W}, \\ d_{m, k} &= c_k e^{j\phi_{m, k}}, \quad e_{m, k} = e^{j\phi_{m, k}}. \end{aligned}$$

The vectors $\mathbf{l}_{n_2}^\sharp$ of size $W = 2M + 1$ are positive combinations of K atoms $\mathbf{a}(f_{n_2, k})$, with $K \leq M$, since we can reasonably assume that the number of lines K is smaller than half the number of pixels M . Thus, Theorem 2 ensures that the decomposition (22) is unique, hence, following (19):

$$(25) \quad \|\mathbf{l}_{n_2}^\sharp\|_A = \sum_{k=1}^K c_k = \hat{\mathbf{x}}^\sharp[0, n_2], \quad \forall n_2 = 0, \dots, H_S - 1 .$$

By contrast, since the coefficients $d_{m, k}$ are complex, Theorem 2 no longer holds and we simply have from Proposition 3:

$$(26) \quad \|\mathbf{t}_m^\sharp\|_A \leq \sum_{k=1}^K c_k, \quad \forall m = -M, \dots, M .$$

Let us take a closer look at the case of one line; that is, $K = 1$, characterized by parameters (θ, η, α) . We recall by (10) that $\hat{\mathbf{x}}^\sharp$ can be written as:

$$\hat{\mathbf{x}}^\sharp[m, n_2] = c_1 e^{j2\pi((f_1 - f_0)n_2 + f_0)m}, \quad c_1 = \frac{\alpha}{\cos \theta}, \quad f_0 = -\frac{\eta}{W}, \quad f_1 = \frac{\tan \theta - \eta}{W} .$$

Let $\mathbf{z} = (z_0, \dots, z_{N-1})$ be a complex vector, whose elements z_i are rearranged in a Toeplitz matrix $\mathbf{P}_K(\mathbf{z})$ of size $(N - K) \times (K + 1)$ and rank K as follows:

$$\mathbf{P}_K(\mathbf{z}) = \begin{pmatrix} z_K & \cdots & z_0 \\ \vdots & \ddots & \vdots \\ z_{N-1} & \cdots & z_{N-K-1} \end{pmatrix} .$$

We get the following characterization of one line in Fourier domain:

PROPOSITION 4. *An image $\hat{\mathbf{x}}$ is of the form $\hat{\mathbf{x}}[m, n] = c_1 e^{j2\pi((f_1 - f_0)n + f_0)m}$ if and only if the columns \mathbf{l}_n and rows \mathbf{t}_m of $\hat{\mathbf{x}}$ are such that $\mathbf{T}_M(\mathbf{l}_n)$ is positive semidefinite and of rank one, $\mathbf{P}_1(\mathbf{t}_m)$ is of rank one and $\hat{\mathbf{x}}[0, n] = \hat{\mathbf{x}}[0, 0]$ for all m and n .*

Proof. See Appendix D. \square

Besides, with $\mathbf{D} = \text{diag}(c_1, \dots, c_K)$ and $\mathbf{V}_{n_2} = [\mathbf{a}(f_{n_2,1}) \ \cdots \ \mathbf{a}(f_{n_2,K})]$, we can remark that

$$\mathbf{T}_M(\mathbf{l}_{n_2}^\#) = \sum_{k=1}^K c_k \mathbf{T}_M(\mathbf{a}(f_{n_2,k})) = \sum_{k=1}^K c_k \mathbf{a}(f_{n_2,k}) \mathbf{a}(f_{n_2,k})^* = \mathbf{V}_{n_2} \mathbf{D} \mathbf{V}_{n_2}^*,$$

where \cdot^* denotes the hermitian conjugate. Since the Toeplitz matrices $\mathbf{T}_M(\mathbf{a}(f_{n_2,k}))$ only contain ones on their main diagonal, then $\frac{1}{M} \text{tr}(\mathbf{T}_M(\mathbf{l}_{n_2}^\#)) = \sum_{k=1}^K c_k$. Moreover, the trace of a positive semidefinite matrix is equal to its nuclear norm, hence:

$$\|\mathbf{T}_M(\mathbf{l}_{n_2}^\#)\|_* = \text{tr}(\mathbf{T}_M(\mathbf{l}_{n_2}^\#)) = M \sum_{k=1}^K c_k = M \|\mathbf{l}_{n_2}^\#\|_{\mathcal{A}}.$$

The nuclear norm of a matrix, which corresponds to the sum of its singular values, is often used as a convex approximation of the rank of this matrix [73, 72]. Consequently, in the following, we consider a convex relaxation of the line characterization given in Proposition 4, in which the rank constraint on $\mathbf{T}_M(\mathbf{l}_{n_2}^\#)$ is replaced by an atomic norm constraint on $\|\mathbf{l}_{n_2}^\#\|_{\mathcal{A}}$. Since the minimum value to achieve is $c^\# = \sum_{k=1}^K c_k$ and since the atomic norm lies on the first row $\hat{\mathbf{x}}[0, n_2] = \hat{\mathbf{x}}[0, 0]$, we impose the constraint $\hat{\mathbf{x}}[0, n_2] = \hat{\mathbf{x}}[0, 0] \leq c^\#$. We do the same for the rows.

REMARK 5. *We showed that if every row and column of the horizontal Fourier transform of the image satisfy assumptions of the Proposition 4, then we have the horizontal Fourier transform of an image containing one line. So, Proposition 4 supports our strategy of dealing with the rows and columns of the image, to reformulate the 2-D problem as a combination of 1-D problems. Proposition 4 shows that, in the case of one line, we do not lose anything by this process; in other words, there is no other image than the image containing one line, whose rows and columns have the prescribed form. This is an indication (not a proof) that the upcoming SDP formulation (27)-(28) is tight: we will not promote structures other than lines in the image. We could not derive a similar characterization for $K \geq 2$ lines, which is a difficult task, but the philosophy remains: the aim is to minimize atomic norms of rows and columns simultaneously, so that the solution will be composed of sparse sums of exponentials in both directions. The convex optimization problem exploiting this strategy is presented in the next section.*

4. Minimization problem with atomic norm regularization. Given the operator $\mathbf{A} : \mathcal{X} \rightarrow \mathcal{Y}$ defined in (14) using the filters (8)-(9) and $\hat{\mathbf{y}}$ the Fourier version of the degraded image observed (15), we are looking for an image $\hat{\mathbf{x}} \in \mathcal{X}$ which minimizes $\|\mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}\|_{\mathcal{Y}}$, for the norm derived from the inner product (32) and whose rows and columns satisfy properties (25) and (26). We fix a constant $c \leq c^\#$. Consequently, the following optimization problem provides an estimator of $\hat{\mathbf{x}}^\#$ defined in (10):

$$(27) \quad \hat{\mathbf{x}} \in \arg \min_{(\hat{\mathbf{x}}, \mathbf{q}) \in \mathcal{X} \times \mathcal{Q}} \frac{1}{2} \|\mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}\|_{\mathcal{Y}}^2,$$

$$\begin{aligned}
(28a) \quad & \widehat{\mathbf{x}}[0, n_2] = \widehat{\mathbf{x}}[0, 0] \leq c, \\
(28b) \quad & \mathbf{q}[m, 0] \leq c, \\
(28c) \quad & \mathbf{T}_{M+1}(\widehat{\mathbf{x}}[:, n_2]) \succcurlyeq 0, \\
(28d) \quad & \mathbf{T}'_{H_S}(\widehat{\mathbf{x}}[m, :], \mathbf{q}[m, :]) \succcurlyeq 0, \\
& \forall n_2 = 0, \dots, H_S - 1, \forall m = 1, \dots, M,
\end{aligned}$$

where the real Hilbert spaces $(\mathcal{X}, \langle \cdot, \cdot \rangle_{\mathcal{X}})$, $(\mathcal{Y}, \langle \cdot, \cdot \rangle_{\mathcal{Y}})$ and $(\mathcal{Q}, \langle \cdot, \cdot \rangle_{\mathcal{Q}})$ are respectively defined by

$$(29) \quad \mathcal{X} = \left\{ \widehat{\mathbf{x}} \in \mathbb{C}^{(M+1) \times H_S} : \text{Im}(\widehat{\mathbf{x}}[0, :]) = 0 \right\},$$

endowed with the following inner product:

$$(30) \quad \langle \widehat{\mathbf{x}}_1, \widehat{\mathbf{x}}_2 \rangle_{\mathcal{X}} = \sum_{n_2=0}^{H_S-1} \widehat{\mathbf{x}}_1[0, n_2] \widehat{\mathbf{x}}_2[0, n_2] + 2\text{Re} \left(\sum_{m=1}^M \sum_{n_2=0}^{H_S-1} \widehat{\mathbf{x}}_1[m, n_2] \widehat{\mathbf{x}}_2[m, n_2]^* \right);$$

$$(31) \quad \mathcal{Y} = \left\{ \widehat{\mathbf{y}} \in \mathbb{C}^{(M+1) \times H} : \text{Im}(\widehat{\mathbf{y}}[0, :]) = 0 \right\},$$

which is equivalent to (29) for $S = 0$, endowed with the inner product:

$$(32) \quad \langle \widehat{\mathbf{y}}_1, \widehat{\mathbf{y}}_2 \rangle_{\mathcal{Y}} = \sum_{n_2=0}^{H-1} \widehat{\mathbf{y}}_1[0, n_2] \widehat{\mathbf{y}}_2[0, n_2] + 2\text{Re} \left(\sum_{m=1}^M \sum_{n_2=0}^{H-1} \widehat{\mathbf{y}}_1[m, n_2] \widehat{\mathbf{y}}_2[m, n_2]^* \right);$$

$$(33) \quad \mathcal{Q} = \left\{ \mathbf{q} \in \mathbb{C}^{(M+1) \times H_S} : \text{Im}(\mathbf{q}[:, 0]) = 0 \right\},$$

endowed with the following inner product:

$$(34) \quad \langle \mathbf{q}_1, \mathbf{q}_2 \rangle_{\mathcal{Q}} = \sum_{m=0}^M \mathbf{q}_1[m, 0] \mathbf{q}_2[m, 0] + 2\text{Re} \left(\sum_{n_2=1}^{H_S-1} \sum_{m=0}^M \mathbf{q}_1[m, n_2] \mathbf{q}_2[m, n_2]^* \right).$$

REMARK 6. *Since we deal with rows and columns of $\widehat{\mathbf{x}} \in \mathcal{X}$ and rows of $\mathbf{q} \in \mathcal{Q}$, we also mention the real Hilbert spaces they belong to, which are respectively denoted by $\mathcal{X}_{\mathbf{l}} \subset \mathbb{R} \times \mathbb{C}^M$, $\mathcal{X}_{\mathbf{t}} \subset \mathbb{C}^{H_S}$ and $\mathcal{Q}_{\mathbf{t}} \subset \mathbb{R} \times \mathbb{C}^{H_S-1}$ endowed with the following inner products:*

$$(35) \quad \langle \mathbf{l}_1, \mathbf{l}_2 \rangle_{\mathcal{X}_{\mathbf{l}}} = \mathbf{l}_1[0] \mathbf{l}_2[0] + 2\text{Re} \left(\sum_{m=1}^M \mathbf{l}_1[m] \mathbf{l}_2[m]^* \right),$$

$$(36) \quad \langle \mathbf{t}_1, \mathbf{t}_2 \rangle_{\mathcal{X}_{\mathbf{t}}} = 2\text{Re} \left(\sum_{n_2=1}^{H_S} \mathbf{t}_1[n_2] \mathbf{t}_2[n_2]^* \right),$$

$$(37) \quad \langle \mathbf{q}_1, \mathbf{q}_2 \rangle_{\mathcal{Q}_{\mathbf{t}}} = \mathbf{q}_1[0] \mathbf{q}_2[0] + 2\text{Re} \left(\sum_{n_2=1}^{H_S-1} \mathbf{q}_1[n_2] \mathbf{q}_2[n_2]^* \right).$$

The operators $\mathbf{T}_{M+1} : \mathcal{X}_{\mathbf{l}} \rightarrow \mathcal{T}_{M+1}$ and $\mathbf{T}'_{H_S} : \mathcal{X}_{\mathbf{t}} \times \mathcal{Q}_{\mathbf{t}} \rightarrow \mathcal{T}_{H_S+1}$ are defined respectively on columns and rows of $\widehat{\mathbf{x}} \in \mathcal{X}$ and $\mathbf{q} \in \mathcal{Q}$, endowed respectively with the

inner products (35) and (36)-(37), to Hermitian Toeplitz matrices of dimension $M+1$ and H_S+1 respectively, whose spaces are denoted by \mathcal{T}_{M+1} and \mathcal{T}_{H_S+1} endowed with the classical inner product on complex matrices:

$$(38) \quad \langle \mathbf{M}, \mathbf{N} \rangle_{\mathcal{M}} = \sum_{i,j} \mathbf{M}_{ij}^* \mathbf{N}_{ij} ,$$

and corresponding Frobenius norm

$$(39) \quad \|\mathbf{M}\|_{\text{F}} = \left(\sum_{i,j} |\mathbf{M}_{ij}|^2 \right)^{1/2} .$$

The expressions of the operators \mathbf{T}_N and \mathbf{T}'_N are given respectively in (20) and (21).

REMARK 7. *This optimization problem could be rewritten in a regularized form involving a parameter λ to tune, what is not any better than the tuning parameter c , which has the advantage of having a physical meaning, related to the line intensities. Moreover, we have a simple estimator of c^\sharp noticing that*

$$\hat{\mathbf{b}}^\sharp[0, n_2] = (\hat{\mathbf{g}}[0] \hat{\mathbf{x}}^\sharp[0, :] * \mathbf{h})[n_2] = \left(\sum_i h[i] \right) \hat{\mathbf{g}}[0] c^\sharp = c^\sharp, \quad \forall n_2 = 1, \dots, H .$$

Then, since $\hat{\mathbf{y}}[0, n_2] = \hat{\mathbf{b}}^\sharp[0, n_2] + \hat{\epsilon}$ with $\mathbb{E}(\hat{\epsilon}) = 0$, one can get an estimation of the parameter c^\sharp by averaging the first row:

$$c \equiv \frac{1}{H} \sum_{n_2=0}^{H-1} \hat{\mathbf{y}}[0, n_2] \approx c^\sharp .$$

We keep this constrained formulation and write it in a more suitable way as follows. Let $\mathcal{H} = \mathcal{X} \times \mathcal{Q}$ be the Hilbert space in which the variable $\mathbf{X} = (\hat{\mathbf{x}}, \mathbf{q})$ lies, endowed with the following inner product:

$$(40) \quad \langle (\hat{\mathbf{x}}_1, \mathbf{q}_1), (\hat{\mathbf{x}}_2, \mathbf{q}_2) \rangle_{\mathcal{H}} = \langle \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 \rangle_{\mathcal{X}} + \langle \mathbf{q}_1, \mathbf{q}_2 \rangle_{\mathcal{Q}} .$$

Let us define $\mathbf{L}_m^{(1)} : \mathcal{H} \rightarrow \mathcal{T}_{H_S+1}$ and $\mathbf{L}_{n_2}^{(2)} : \mathcal{H} \rightarrow \mathcal{T}_{M+1}$ by

$$(41) \quad \mathbf{L}_m^{(1)}(\mathbf{X}) = \mathbf{T}'_{H_S}(\hat{\mathbf{x}}[m, :], \mathbf{q}[m, :]) ,$$

$$(42) \quad \mathbf{L}_{n_2}^{(2)}(\mathbf{X}) = \mathbf{T}_{M+1}(\hat{\mathbf{x}}[:, n_2]) .$$

We denote by ι_C the indicator function of a convex set C , defined by

$$\iota_C : x \mapsto \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases} .$$

We denote by \mathcal{C} the cone of positive semidefinite matrices and we introduce the set $\mathcal{B} \subset \mathcal{H}$ corresponding to the boundary constraints:

$$(43) \quad \mathcal{B} = \left\{ (\hat{\mathbf{x}}, \mathbf{q}) \in \mathcal{H} : \hat{\mathbf{x}}[0, n_2] = \hat{\mathbf{x}}[0, 0] \leq c, \mathbf{q}[m, 0] \leq c \right\} .$$

Then the optimization problem (27) under constraints (28) can be rewritten as follows:
(44)

$$\tilde{\mathbf{X}} = \arg \min_{\mathbf{X}=(\hat{\mathbf{x}}, \mathbf{q}) \in \mathcal{H}} \left\{ \frac{1}{2} \|\mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}\|_{\mathcal{Y}}^2 + \iota_{\mathcal{B}}(\mathbf{X}) + \sum_{m=1}^M \iota_{\mathcal{C}}(\mathbf{L}_m^{(1)}(\mathbf{X})) + \sum_{n_2=0}^{H_S-1} \iota_{\mathcal{C}}(\mathbf{L}_{n_2}^{(2)}(\mathbf{X})) \right\}.$$

We now propose two different algorithms to solve this convex optimization problem. The first one in subsection 4.1 is more general and can be applied to the extended setting (with no restriction on the line angles) and inpainting problems, presented in subsection 4.3 and subsection 4.4 respectively, whereas the second cannot, since it is difficult to compute the proximity operator associated to the data fidelity term. However, the second algorithm detailed in subsection 4.2, happens to be faster and so will be used for most of the numerical experiments exposed in section 7.

4.1. First algorithm design. The optimization problem (44) can be cast as a minimization problem, involving smooth, proximable and linear composite terms [23]:

$$(45) \quad \tilde{\mathbf{X}} = \arg \min_{\mathbf{X} \in \mathcal{H}} \left\{ F(\mathbf{X}) + G(\mathbf{X}) + \sum_{i=0}^{Q-1} H_i(\mathbf{L}_i(\mathbf{X})) \right\},$$

with $F(\mathbf{X}) = \frac{1}{2} \|\mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}\|_{\mathcal{Y}}^2$, $\mathbf{X} = (\hat{\mathbf{x}}, \mathbf{q})$, $G = \iota_{\mathcal{B}}$, which is proximable, $Q = M + H_S$ linear composite terms where $H_i = \iota_{\mathcal{C}}$, $\mathbf{L}_i = \mathbf{L}_i^{(2)}$ when $0 \leq i \leq H_S - 1$ and $\mathbf{L}_i = \mathbf{L}_{i-H_S+1}^{(1)}$ when $H_S \leq i \leq H_S + M - 1$. We define $\mathbf{H}\mathbf{x} = \sum_{i=0}^{Q-1} H_i x_i$, $\mathbf{L}^{(1)}(\mathbf{X}) = (\mathbf{L}_1^{(1)}(\mathbf{X}), \dots, \mathbf{L}_M^{(1)}(\mathbf{X}))$ and $\mathbf{L}^{(2)}(\mathbf{X}) = (\mathbf{L}_0^{(2)}(\mathbf{X}), \dots, \mathbf{L}_{H_S-1}^{(2)}(\mathbf{X}))$. $\mathbf{L} = (\mathbf{L}^{(1)}, \mathbf{L}^{(2)})$ is the linear operator such that the composite terms can be written $\mathbf{H} \circ \mathbf{L}$, where \circ denotes composition. We define an inner product on the range of $\mathbf{L}^{(1)}$, $\mathbf{L}^{(2)}$ and \mathbf{L} , which are product spaces, as the sum of the inner products defined on these spaces (similarly to (40)). We denote the corresponding norms by $\|\cdot\|_{(1)}$, $\|\cdot\|_{(2)}$ and $\|\cdot\|_{(1,2)}$. We define the following operator norms:

$$(46) \quad \|\mathbf{A}\| = \sup_{\hat{\mathbf{x}} \in \mathcal{X}} \frac{\|\mathbf{A}\hat{\mathbf{x}}\|_{\mathcal{Y}}}{\|\hat{\mathbf{x}}\|_{\mathcal{X}}},$$

$$(47) \quad \|\mathbf{L}_i\| = \sup_{\mathbf{X} \in \mathcal{H}} \frac{\|\mathbf{L}_i(\mathbf{X})\|_{\mathcal{F}}}{\|\mathbf{X}\|_{\mathcal{H}}},$$

$$(48) \quad \|\mathbf{L}^{(j)}\| = \sup_{\mathbf{X} \in \mathcal{H}} \frac{\|\mathbf{L}^{(j)}(\mathbf{X})\|_{(j)}}{\|\mathbf{X}\|_{\mathcal{H}}}, \quad j \in \{1, 2\},$$

$$(49) \quad \|\mathbf{L}\| = \sup_{\mathbf{X} \in \mathcal{H}} \frac{\|\mathbf{L}(\mathbf{X})\|_{(1,2)}}{\|\mathbf{X}\|_{\mathcal{H}}}.$$

We now establish some properties of these functions, operators and norms.

LEMMA 5. The norm (46) of the operator \mathbf{A} defined in (14) is given by

$$(50) \quad \|\mathbf{A}\| = \|\hat{\mathbf{g}}\|_{\infty} \|\hat{\mathbf{h}}\|_{\infty}.$$

Proof. If we denote by $\hat{\mathbf{x}}_k$ the $(k+1)$ -th row of $\hat{\mathbf{x}}$, then by definition the operator \mathbf{A} maps the $(k+1)$ -th row of $\hat{\mathbf{x}}$ to $\hat{\mathbf{g}}[k](\hat{\mathbf{x}}_k * \mathbf{h})$. Let us calculate the norm of this operator. By considering in Fourier the norm operator $f \mapsto f * h$ we have the inequality

$$\|\hat{\mathbf{x}}_k * \mathbf{h}\|_2 \leq \|\hat{\mathbf{h}}\|_{\infty} \|\hat{\mathbf{x}}_k\|_2.$$

Thus, with the norm derived from (30) we get

$$\begin{aligned}
 (51) \quad \|\mathbf{A}\hat{\mathbf{x}}\|_{\mathbf{y}}^2 &= |\hat{g}_0|^2 \|\hat{\mathbf{x}}_0 * \mathbf{h}\|_2^2 + 2|\hat{g}_1|^2 \|\hat{\mathbf{x}}_1 * \mathbf{h}\|_2^2 + \cdots + 2|\hat{g}_M|^2 \|\hat{\mathbf{x}}_M * \mathbf{h}\|_2^2, \\
 &\leq \|\hat{\mathbf{g}}\|_{\infty}^2 \|\hat{\mathbf{h}}\|_{\infty}^2 (\|\hat{\mathbf{x}}_0\|_2^2 + 2\|\hat{\mathbf{x}}_1\|_2^2 + \cdots + 2\|\hat{\mathbf{x}}_M\|_2^2), \\
 &\leq \|\hat{\mathbf{g}}\|_{\infty}^2 \|\hat{\mathbf{h}}\|_{\infty}^2 \|\hat{\mathbf{x}}\|_{\mathcal{X}}^2.
 \end{aligned}$$

Since the filter \mathbf{h} is lowpass, the equality is attained for an image $\hat{\mathbf{x}}$ whose all rows are zero, except one which is constant (and nonzero), of index m_0 , where \hat{g}_{m_0} corresponds to the maximum $\|\hat{\mathbf{g}}\|_{\infty}$, which proves the result (50). \square

LEMMA 6. *The adjoint operator of the operator \mathbf{A} is denoted by \mathbf{A}^* and defined such that $\langle \mathbf{A}\hat{\mathbf{x}}, \hat{\mathbf{z}} \rangle_{\mathbf{y}} = \langle \hat{\mathbf{x}}, \mathbf{A}^*\hat{\mathbf{z}} \rangle_{\mathcal{X}}$. Its matrix expression is given by*

$$(52) \quad \mathbf{A}^*\hat{\mathbf{z}} = \hat{\mathbf{G}}^*\hat{\mathbf{z}}\check{\mathbf{H}}.$$

Proof. From the definition (14), we have the matrix product $\mathbf{A}\hat{\mathbf{x}} = \hat{\mathbf{G}}\hat{\mathbf{x}}\check{\mathbf{H}}^{\top}$. Then we use the fact that for any matrix \mathbf{M} , we have $\langle \mathbf{M}\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 \rangle_{\mathcal{M}} = \langle \hat{\mathbf{x}}_1, \mathbf{M}^*\hat{\mathbf{x}}_2 \rangle_{\mathcal{M}}$ and we remark that

$$\langle \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 \rangle_{\mathcal{X}} = \langle \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 \rangle_{\mathcal{M}} + \langle \hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2 \rangle_{\mathcal{M}}^* - \langle \hat{\mathbf{x}}_1[0, :], \hat{\mathbf{x}}_2[0, :] \rangle_{\mathbb{R}^{H_S}}.$$

This yields the claimed result. \square

Then, we have the following proposition:

PROPOSITION 7. *For $\mathbf{X} = (\hat{\mathbf{x}}, \mathbf{q})$, the gradient of $F(\mathbf{X}) = \frac{1}{2} \|\mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}\|_{\mathbf{y}}^2$ is*

$$\nabla F(\mathbf{X}) = (\mathbf{A}^*(\mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}), \mathbf{0})^{\top},$$

which is Lipschitz-continuous with Lipschitz constant $\beta = \|\hat{\mathbf{g}}\|_{\infty}^2 \|\hat{\mathbf{h}}\|_{\infty}^2$.

Proof. See in Appendix G. \square

We now give in the next proposition the adjoint operators of:

$$(53) \quad \mathbf{T}_{M+1} : (\mathcal{X}_t, \langle \cdot, \cdot \rangle_{\mathcal{X}_t}) \rightarrow (\mathcal{T}_{M+1}, \langle \cdot, \cdot \rangle_{\mathcal{M}}),$$

$$(54) \quad \mathbf{T}'_{H_S} : (\mathcal{X}_t \times \mathcal{Q}_t, \langle \cdot, \cdot \rangle_{\mathcal{X}_t \times \mathcal{Q}_t}) \rightarrow (\mathcal{T}_{H_S+1}, \langle \cdot, \cdot \rangle_{\mathcal{M}}),$$

where the inner products are defined in (35), (36), (37) and (38).

PROPOSITION 8. *For $\mathbf{M}^{(1)} \in \mathcal{T}_{H_S+1}$ and $\mathbf{M}^{(2)} \in \mathcal{T}_{M+1}$ the adjoint operators of (53) and (54) applied to $\mathbf{M}^{(1)}$ and $\mathbf{M}^{(2)}$ give the vectors*

$$\begin{aligned}
 \mathbf{z}_2 &= \mathbf{T}_{M+1}^* \mathbf{M}^{(2)} \in \mathbb{R} \times \mathbb{C}^M, \\
 (\mathbf{z}_1, \mathbf{q}_1) &= \mathbf{T}'_{H_S}{}^* \mathbf{M}^{(1)} \in \mathbb{C}^{H_S} \times (\mathbb{R} \times \mathbb{C}^{H_S-1}),
 \end{aligned}$$

respectively, whose components are:

$$\begin{aligned}
 \mathbf{z}_2[k] &= \sum_{l=0}^{M-k} \mathbf{M}_{l+k,l}^{(2)}, \quad \forall k = 0, \dots, M, \\
 \mathbf{z}_1[k] &= \mathbf{M}_{H_S+1,k}^{(1)}, \quad \mathbf{q}_1[k] = \sum_{l=0}^{H_S-1-k} \mathbf{M}_{l+k,l}^{(1)} + \delta_k \mathbf{M}_{H_S,H_S}^{(1)}, \quad \forall k = 0, \dots, H_S - 1.
 \end{aligned}$$

Proof. See in [Appendix E](#). □

Now we will provide an explicit upper bound for the operator norm $\|\mathbf{L}\|$.

PROPOSITION 9. *The norm of the operator $\mathbf{L} = (\mathbf{L}^{(1)}, \mathbf{L}^{(2)})$, where*

$$\begin{aligned}\mathbf{L}^{(1)}(\mathbf{X}) &= (\mathbf{L}_1^{(1)}(\mathbf{X}), \dots, \mathbf{L}_M^{(1)}(\mathbf{X})) , \\ \mathbf{L}^{(2)}(\mathbf{X}) &= (\mathbf{L}_0^{(2)}(\mathbf{X}), \dots, \mathbf{L}_{H_S-1}^{(2)}(\mathbf{X})) ,\end{aligned}$$

with $\mathbf{L}_m^{(1)}$ and $\mathbf{L}_{n_2}^{(2)}$ defined in (41)-(42), satisfies

$$\|\mathbf{L}\|^2 \leq \|\mathbf{L}^{(1)}\|^2 + \|\mathbf{L}^{(2)}\|^2 = \|\mathbf{T}'_{H_S}\|^2 + \|\mathbf{T}_{M+1}\|^2 = (H_S + 1) + (M + 1) .$$

Proof. See in [Appendix F](#). □

To solve the problem (44), we first propose [Algorithm 1](#), which uses the primal-dual method introduced in [66]. Following [23, Theorem 5.1], we know that the method converges to a solution $(\tilde{\mathbf{X}}, \tilde{\xi}_0, \dots, \tilde{\xi}_{Q-1})$ of the problem (45), provided the parameters $\tau > 0$ and $\sigma > 0$ in [Algorithm 1](#) are such that

$$(55) \quad \frac{1}{\tau} - \sigma \|\mathbf{L}\|^2 > \frac{\beta}{2} .$$

We then choose $0 < \tau < 2$, $\sigma = (H_S + M + 2)^{-2}(1/\tau - \beta/1.9)$ and $\rho_n \equiv \rho = 1$.

Algorithm 1 Primal-dual splitting algorithm for (45)

Input: $\hat{\mathbf{y}}$ 1-D FFT of the blurred and noisy data image \mathbf{y}

Output: $\tilde{\mathbf{x}}$ solution of the optimization problem (27) under constraints (28)

- 1: Initialize primal and dual variables to zero $\mathbf{X}_0 = 0$, $\xi_{i,0} = 0$, $\forall i \in \llbracket 1, Q \rrbracket$
 - 2: **for** $n = 1$ **to** Number of iterations **do**
 - 3: $\mathbf{X}_{n+1} = \text{prox}_{\tau G}(\mathbf{X}_n - \tau \nabla F(\mathbf{X}_n) - \tau \sum_{i=0}^{Q-1} \mathbf{L}_i^* \xi_{i,n})$,
 - 4: $\mathbf{X}_{n+1} = \rho_n \tilde{\mathbf{X}}_{n+1} - (1 - \rho_n) \mathbf{X}_n$,
 - 5: **for** $i = 0$ **to** $Q - 1$ **do**
 - 6: $\xi_{i,n+1} = \text{prox}_{\sigma H_i^*}(\xi_{i,n} + \sigma \mathbf{L}_i(2\mathbf{X}_{n+1} - \mathbf{X}_n))$,
 - 7: $\xi_{i,n+1} = \rho_n \tilde{\xi}_{i,n+1} + (1 - \rho_n) \xi_{i,n}$,
 - 8: **end for**
 - 9: **end for**
-

We detail below the other terms in lines 3 and 6 of [Algorithm 1](#), involving the computation of proximity operators and adjoint operators. For more details on convex analysis, monotone operator theory and proximal splitting methods, we refer the reader to [8, 3, 22, 62].

Set $\overline{x_0} = \frac{1}{H_S} \sum_{n_2=0}^{H_S-1} \hat{\mathbf{x}}[0, n_2]$, set $G = \iota_{\mathcal{B}}$, with \mathcal{B} defined in (43). Then we have, for every m, n_2 :

$$\text{prox}_{\tau G}(\hat{\mathbf{x}}, \mathbf{q}) = \begin{cases} \hat{\mathbf{x}}[0, n_2] = \overline{x_0} & \text{if } \overline{x_0} \leq c \\ \hat{\mathbf{x}}[0, n_2] = c & \text{otherwise} \\ \mathbf{q}[m, 0] = c & \text{if } \mathbf{q}[m, 0] > c \end{cases} .$$

Let $P_{\mathcal{C}}$ be the projection operator onto the cone of positive matrix \mathcal{C} ; by Moreau

identity [3],

$$\text{prox}_{\sigma H_i^*}(\mathbf{M}) = \mathbf{M} - \sigma \text{prox}_{\frac{H_i}{\sigma}}\left(\frac{1}{\sigma}\mathbf{M}\right) = \mathbf{M} - P_{\mathcal{C}}(\mathbf{M}) .$$

Finally, we need to compute, in line 3 of the algorithm, the adjoint operators L_i^* , where the operators L_i are defined in (41)-(42). The dual variables $(\xi_{i,n})_i$ in Algorithm 1 refer to Hermitian Toeplitz matrices $\mathbf{M}_m^{(1)} \in \mathcal{T}_{H_S+1}$ or $\mathbf{M}_{n_2}^{(2)} \in \mathcal{T}_{M+1}$. By definition, the adjoint operators give the images $(\mathbf{z}_m^{(1)}, \mathbf{q}_m^{(1)}) = L_m^{(1)*} \mathbf{M}_m^{(1)}$ and $(\mathbf{z}_{n_2}^{(2)}, \mathbf{q}_{n_2}^{(2)}) = L_{n_2}^{(2)*} \mathbf{M}_{n_2}^{(2)}$. According to the definitions (41)-(42), for a primal variable $\mathbf{X} = (\hat{\mathbf{x}}, \mathbf{q})$ the operators $L_m^{(1)}$ and $L_{n_2}^{(1)}$ act respectively on the $(m+1)$ -th row of the images $(\hat{\mathbf{x}}, \mathbf{q})$ and on the (n_2+1) -th column of the image $\hat{\mathbf{x}}$ only; so we can easily see, concerning the adjoint operators, that $\mathbf{q}_{n_2}^{(2)} = \mathbf{0}$ and $\mathbf{z}_{n_2}^{(2)}$ (resp. $\mathbf{z}_m^{(1)}, \mathbf{q}_m^{(1)}$) is zero, except at the corresponding column index n_2 (resp. row index m), where

$$(56) \quad \mathbf{z}_{n_2}^{(2)}[:, n_2] = \mathbf{T}_{M+1}^* \mathbf{M}_{n_2}^{(2)} ,$$

$$(57) \quad (\mathbf{z}_m^{(1)}[m, :], \mathbf{q}_m^{(1)}[m, :]) = \mathbf{T}_{H_S}'^* \mathbf{M}_m^{(1)} ,$$

with the expression of the adjoint operators \mathbf{T}_{M+1}^* and $\mathbf{T}_{H_S}'^*$ given in Proposition 8. Thus, the operations on $\mathbf{X}_n = (\hat{\mathbf{x}}_n, \mathbf{q}_n)$ before applying $\text{prox}_{\tau G}$ consist in a gradient descent step $\mathbf{X}_n - \tau \nabla F(\mathbf{X}_n)$, followed by an update of all its rows and columns due to the terms $-\tau \sum_{i=0}^{Q-1} L_i^* \xi_{i,n}$, whose expressions are provided by (56) and (57).

4.2. Second algorithm design. We can note that in Algorithm 1, τ must be smaller than $2/\beta$, which is a limitation in terms of convergence speed. To overcome this issue, we subsequently developed a second algorithm, similar to Algorithm 1, but with the data fidelity term $\|\mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}\|_{\mathcal{Y}}$ activated through its proximity operator, instead of its gradient. We consider solving the optimization problem by an over-relaxed version [23] of the Chambolle–Pock algorithm [17]:

$$(58) \quad \mathbf{X}^* = \arg \min_{\mathbf{X} \in \mathcal{H}} \{G(\mathbf{X}) + \mathbf{H}(\mathbf{L}(\mathbf{X}))\} ,$$

with now $G = \frac{1}{2} \|\mathbf{A} \cdot -\hat{\mathbf{y}}\|_{\mathcal{Y}}^2$ which is proximable, $\mathbf{H}\mathbf{x} = \sum_{i=0}^Q H_i x_i$ with $H_i = \iota_{\mathcal{C}}$ for $i < Q$, where $L_i = L_i^{(2)}$ when $0 \leq i \leq H_S - 1$, $L_i = L_{i-H_S+1}^{(1)}$ when $H_S \leq i \leq Q - 1$ and $H_Q = \iota_{\mathcal{B}}$ with $L_Q = \text{Id}$. So now, $\|\mathbf{L}\|^2 \leq H_S + M + 3$.

Let $\tau > 0$ and $\sigma > 0$ such that $\tau\sigma\|\mathbf{L}\|^2 = 1$, then the primal-dual Algorithm 2, with $F = 0$ and weights $\rho_n \equiv \rho = 1.9$, which is an over-relaxed version of the Chambolle–Pock algorithm, converges to a solution $(\tilde{\mathbf{X}}, \tilde{\xi}_0, \dots, \tilde{\xi}_{Q-1})$ of the problem (45) [23, Theorem 5.1].

The Algorithm 2 requires computing $\text{prox}_{\tau G}$. Since we have

$$(59) \quad \begin{aligned} \mathbf{p} = \text{prox}_{\tau G}(\hat{\mathbf{x}}) &\Leftrightarrow \hat{\mathbf{x}} - \mathbf{p} = \nabla(\tau G)(\mathbf{p}) , \\ &\Leftrightarrow \hat{\mathbf{x}} - \mathbf{p} = \tau \mathbf{A}^* (\mathbf{A}\mathbf{p} - \hat{\mathbf{y}}) , \\ &\Leftrightarrow \hat{\mathbf{x}} + \tau \mathbf{A}^* \hat{\mathbf{y}} = (\mathbf{I} + \tau \mathbf{A}^* \mathbf{A}) \mathbf{p} , \end{aligned}$$

the proximity operator has the following expression:

$$\text{prox}_{\tau G}(\hat{\mathbf{x}}) = (\mathbf{I} + \tau \mathbf{A}^* \mathbf{A})^{-1} (\hat{\mathbf{x}} + \tau \mathbf{A}^* \hat{\mathbf{y}}) ,$$

Algorithm 2 Primal-dual splitting algorithm for (58)**Input:** $\hat{\mathbf{y}}$ 1-D FFT of the blurred and noisy data image \mathbf{y} **Output:** $\tilde{\mathbf{x}}$ solution of the optimization problem (27) under constraints (28)

- 1: Initialize all primal and dual variables to zero
- 2: **for** $n = 1$ **to** Number of iterations **do**
- 3: $\tilde{\mathbf{X}}_{n+1} = \text{prox}_{\tau G}(\mathbf{X}_n - \tau \sum_{i=0}^{Q-1} \mathbf{L}_i^* \boldsymbol{\xi}_{i,n}),$
- 4: $\mathbf{X}_{n+1} = \rho_n \tilde{\mathbf{X}}_{n+1} - (1 - \rho_n) \mathbf{X}_n$
- 5: **for** $i = 0$ **to** $Q - 1$ **do**
- 6: $\tilde{\boldsymbol{\xi}}_{i,n+1} = \text{prox}_{\sigma H_i^*}(\boldsymbol{\xi}_{i,n} + \sigma \mathbf{L}_i(2\mathbf{X}_{n+1} - \mathbf{X}_n)),$
- 7: $\boldsymbol{\xi}_{i,n+1} = \rho_n \tilde{\boldsymbol{\xi}}_{i,n+1} + (1 - \rho_n) \boldsymbol{\xi}_{i,n}$
- 8: **end for**
- 9: **end for**

for which we propose below two ways of computing the inverse.

We proved in Lemma 6 that $\mathbf{A}^* \hat{\mathbf{y}} = \hat{\mathbf{G}}^* \hat{\mathbf{y}} \check{\mathbf{H}}$ and then

$$(\mathbf{I} + \tau \mathbf{A}^* \mathbf{A}) \hat{\mathbf{x}} = \hat{\mathbf{x}} + \mathbf{P} \hat{\mathbf{x}} \mathbf{Q}, \quad \mathbf{P} = \tau \mathbf{G}^* \mathbf{G}, \quad \mathbf{Q} = \check{\mathbf{H}}^T \check{\mathbf{H}}.$$

The square matrices \mathbf{P} and \mathbf{Q} are of size $p = M + 1$ and $q = H_S$. We have to solve $(\mathbf{I} + \tau \mathbf{A}^* \mathbf{A}) \hat{\mathbf{x}} = \mathbf{z}$; that is, $\hat{\mathbf{x}} + \mathbf{P} \hat{\mathbf{x}} \mathbf{Q} = \mathbf{z}$. This kind of system can be solved by the mean of the Kronecker Product (2) as:

$$\hat{\mathbf{x}} + \mathbf{P} \hat{\mathbf{x}} \mathbf{Q} = \mathbf{z} \iff (\mathbf{I}_{pq} + \mathbf{Q} \otimes \mathbf{P}^T) \text{vec}(\hat{\mathbf{x}}) = \text{vec}(\mathbf{z}).$$

where $\text{vec}(\hat{\mathbf{x}})$ denotes the vectorization of the matrix $\hat{\mathbf{x}}$ formed by stacking the columns of $\hat{\mathbf{x}}$ into a single column vector and $\mathbf{I}_{pq} + \mathbf{Q} \otimes \mathbf{P}^T$ is a matrix of size $pq \times pq$ which can be inverted, giving access to $\text{vec}(\hat{\mathbf{x}})$ and then to $\hat{\mathbf{x}}$. Finally, the operator $\text{prox}_{\tau G}$ can be seen as a large matrix-vector product.

Another option consists in operating on the rows $\hat{\mathbf{x}}_m$ of $\hat{\mathbf{x}}$, since the operator \mathbf{A} acts on them:

$$(\mathbf{I} + \tau \mathbf{A}^* \mathbf{A}) \hat{\mathbf{x}} = \mathbf{z} \iff (\mathbf{I} + |\hat{\mathbf{g}}_m|^2 \mathbf{Q}) \hat{\mathbf{x}}_m = \mathbf{z}_m, \quad \forall m = 0, \dots, M.$$

This time, the operator $\text{prox}_{\tau G}$ involves performing $M + 1$ matrix-vector products of size $q \times q$, which appears to be more efficient in practice.

4.3. Extended problem formulation. We now consider a data image \mathbf{b}^\sharp containing lines with no angle restriction, which extends the previous case by relaxing the assumption made in Hypothesis 1. We can decompose this image into the sum of two images $\mathbf{b}^\sharp = \mathbf{b}_1^\sharp + \mathbf{b}_2^\sharp$, with \mathbf{b}_1^\sharp (resp. \mathbf{b}_2^\sharp) containing vertical (resp. horizontal) lines, that is with angles in $(-\pi/4, \pi/4]$ (resp. outside this range). We can also define $\hat{\mathbf{x}}_1^\sharp$ of size $(M + 1) \times H_S$ and $\hat{\mathbf{x}}_2^\sharp$ of size $W_S \times (P + 1)$ with $W_S = W + 2S$ and $P = (H - 1)/2$ such as $\mathbf{A} \hat{\mathbf{x}}_1^\sharp = \hat{\mathbf{b}}_1^\sharp$ and $\tilde{\mathbf{A}} \hat{\mathbf{x}}_2^\sharp = \hat{\mathbf{b}}_2^\sharp$, where $\mathbf{g}_2 = (\mathbf{0}_{P-S}, \mathbf{h}, \mathbf{0}_{P-S})$ and $\tilde{\mathbf{A}}$ denotes the operator which multiplies each column vector $\hat{\mathbf{x}}_2^\sharp[:, n_2]$ by the corresponding Fourier coefficient $\hat{\mathbf{g}}_2[n_2]$ and convolves it with the filter \mathbf{h} ; that is, $\tilde{\mathbf{A}} \hat{\mathbf{x}}_2 = (\hat{\mathbf{G}}_2 \hat{\mathbf{x}}_2) * \mathbf{h}$ with $\hat{\mathbf{G}}_2 = \text{diag}(\hat{\mathbf{g}}_2[0], \dots, \hat{\mathbf{g}}_2[P])$. We finally define the Hermitian symmetry operator \mathbf{S}_1 (resp. \mathbf{S}_2), which to each column $\mathbf{v} = (v_0, v_1, \dots, v_M)$ (resp. row $\mathbf{v} = (v_0, v_1, \dots, v_P)$) associates the symmetric extension $(v_M^*, \dots, v_0, \dots, v_M)$ (resp. $(v_P^*, \dots, v_0, \dots, v_P)$). Let $\mathbf{X}_1 = (\hat{\mathbf{x}}_1, \mathbf{q}_1)$ and $\mathbf{X}_2 = (\hat{\mathbf{x}}_2, \mathbf{q}_2)$ be the optimization variables, living in spaces

$\mathcal{H}_1 = \mathcal{X}_1 \times \mathcal{Q}_1$ and $\mathcal{H}_2 = \mathcal{X}_2 \times \mathcal{Q}_2$. Let $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$, $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$ and $\mathcal{Q} = \mathcal{Q}_1 \times \mathcal{Q}_2$. The data fidelity term is now:

$$F(\mathbf{X}_1, \mathbf{X}_2) = \frac{1}{2} \|\mathcal{F}_1^{-1} \mathbf{S}_1 \mathbf{A} \hat{\mathbf{x}}_1 + \mathcal{F}_2^{-1} \mathbf{S}_2 \tilde{\mathbf{A}} \hat{\mathbf{x}}_2 - \mathbf{y}\|_{\mathbb{F}}^2 = \frac{1}{2} \|\mathbf{A}_1 \hat{\mathbf{x}}_1 + \mathbf{A}_2 \hat{\mathbf{x}}_2 - \mathbf{y}\|_{\mathbb{F}}^2,$$

with $\mathbf{A}_1 = \mathcal{F}_1^{-1} \mathbf{S}_1 \mathbf{A}$, $\mathbf{A}_2 = \mathcal{F}_2^{-1} \mathbf{S}_2 \tilde{\mathbf{A}}$, where \mathcal{F}_1 (resp. \mathcal{F}_2) is the Fourier transform with respect to the columns (resp. rows) and $\|\cdot\|_{\mathbb{F}}$ is the Frobenius norm.

PROPOSITION 10. *The gradient of F is*

$$(60) \quad \nabla F(\mathbf{X}_1, \mathbf{X}_2) = \frac{1}{2} \begin{pmatrix} \mathbf{A}_1^* (\mathbf{A}_1 \hat{\mathbf{x}}_1 + \mathbf{A}_2 \hat{\mathbf{x}}_2 - \mathbf{y}) \\ \mathbf{A}_2^* (\mathbf{A}_1 \hat{\mathbf{x}}_1 + \mathbf{A}_2 \hat{\mathbf{x}}_2 - \mathbf{y}) \end{pmatrix},$$

which is Lipschitz-continuous of Lipschitz constant $\beta = \frac{1}{\min(W, H)}$.

Proof. See in [Appendix H](#). □

The image $\hat{\mathbf{x}}_1^\#$ keeps the same kind of constraints as in the [Algorithm 1](#), which act similarly on the image $\hat{\mathbf{x}}_2^\#$ in a rotated way; that is, we define

$$(61) \quad \mathbf{L}_m^{(3)}(\mathbf{X}_2) = \mathbf{T}_{P+1}(\hat{\mathbf{x}}_2[m, :]),$$

$$(62) \quad \mathbf{L}_{n_2}^{(4)}(\mathbf{X}_2) = \mathbf{T}'_{W_S}(\text{fliplr}(\hat{\mathbf{x}}_2[:, n_2]), \text{fliplr}(\mathbf{q}_2[:, n_2])),$$

where `fliplr` denotes a flip from left to right on each column of the matrix. The boundary constraints on $\hat{\mathbf{x}}_1$ and $\hat{\mathbf{x}}_2$ are respectively given by:

$$(63) \quad \mathcal{B}_1 = \left\{ (\hat{\mathbf{x}}_1, \mathbf{q}_1) \in \mathcal{H}_1 : \hat{\mathbf{x}}_1[0, n_2] = \hat{\mathbf{x}}_1[0, 0] \leq c_1, \mathbf{q}_1[m, 0] \leq c_1 \right\},$$

$$(64) \quad \mathcal{B}_2 = \left\{ (\hat{\mathbf{x}}_2, \mathbf{q}_2) \in \mathcal{H}_2 : \hat{\mathbf{x}}_2[m, 0] = \hat{\mathbf{x}}_2[0, 0] \leq c_2, \mathbf{q}_2[P, n_2] \leq c_2 \right\}.$$

Likewise, the inner product on spaces \mathcal{X}_2 and \mathcal{Q}_2 are:

$$(65) \quad \langle \mathbf{z}_1, \mathbf{z}_2 \rangle_{\mathcal{X}_2} = \sum_{m=0}^{W_S-1} \mathbf{z}_1[m, 0] \mathbf{z}_2[m, 0] + 2\text{Re} \left(\sum_{n_2=1}^P \sum_{m=0}^{W_S-1} \mathbf{z}_1[m, n_2] \mathbf{z}_2[m, n_2]^* \right),$$

$$(66) \quad \langle \mathbf{z}_1, \mathbf{z}_2 \rangle_{\mathcal{Q}_2} = 2\text{Re} \left(\sum_{n_2=0}^P \sum_{m=0}^{W_S-1} \mathbf{z}_1[m, n_2] \mathbf{z}_2[m, n_2]^* \right),$$

and so the adjoint of the operators remain the same.

Like before, we define $\mathbf{L}^{(3)}(\mathbf{X}_2) = (\mathbf{L}_0^{(3)}(\mathbf{X}_2), \dots, \mathbf{L}_{W_S-1}^{(3)}(\mathbf{X}_2))$, also $\mathbf{L}^{(4)}(\mathbf{X}_2) = (\mathbf{L}_0^{(4)}(\mathbf{X}_2), \dots, \mathbf{L}_P^{(4)}(\mathbf{X}_2))$ and $\mathbf{L} = (\mathbf{L}^{(1)}, \mathbf{L}^{(2)}, \mathbf{L}^{(3)}, \mathbf{L}^{(4)})$. It is easy to show that

$$\begin{aligned} \|\mathbf{L}\|^2 &\leq \|\mathbf{L}^{(1)}\|^2 + \|\mathbf{L}^{(2)}\|^2 + \|\mathbf{L}^{(3)}\|^2 + \|\mathbf{L}^{(4)}\|^2, \\ &\leq (H_S + 1) + (M + 1) + (P + 1) + (W_S + 1). \end{aligned}$$

Finally, we have

$$(67) \quad (\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2) = \arg \min_{(\mathbf{X}_1, \mathbf{X}_2) \in \mathcal{H}} \left\{ \frac{1}{2} \|\mathbf{A}_1 \hat{\mathbf{x}}_1 + \mathbf{A}_2 \hat{\mathbf{x}}_2 - \mathbf{y}\|_{\mathbb{F}}^2 \right. \\ \left. + \iota_{\mathcal{B}_1}(\mathbf{X}_1) + \sum_{m=1}^M \iota_{\mathcal{C}}(\mathbf{L}_m^{(1)}(\mathbf{X}_1)) + \sum_{n_2=0}^{H_S-1} \iota_{\mathcal{C}}(\mathbf{L}_{n_2}^{(2)}(\mathbf{X}_1)) \right. \\ \left. + \iota_{\mathcal{B}_2}(\mathbf{X}_2) + \sum_{m=0}^{W_S-1} \iota_{\mathcal{C}}(\mathbf{L}_m^{(3)}(\mathbf{X}_2)) + \sum_{n_2=1}^P \iota_{\mathcal{C}}(\mathbf{L}_{n_2}^{(4)}(\mathbf{X}_2)) \right\}.$$

4.4. Inpainting problems. We now consider the case in which a binary mask is applied on the data image, like in [Figure 3](#). The corresponding linear operator, denoted by \mathbf{M} , consists in elementwise multiplication of the matrix $\mathbf{b}^\#$ with a binary matrix, whose zero coefficients are the indices of the pixels unavailable to observation. We have $\mathbf{M}^* = \mathbf{M}$. The data fidelity term becomes $F(\mathbf{X}) = \frac{1}{2} \|\mathbf{M} \mathcal{F}_1^{-1} \mathbf{S}_1 \mathbf{A} \hat{\mathbf{x}} - \mathbf{y}\|_{\mathbb{F}}^2$, whose gradient can be expressed as previously, with $\beta = 1/W$ (since $\|\mathcal{F}_1^{-1}\| = 1/W$ and $\|\mathbf{M}\| = 1$). The constraints remain the same as in [\(28\)](#) and the method is also easily transposable to the extended setting of [subsection 4.3](#).

At this point, the first part of the process has been completed; that is, the image $\hat{\mathbf{x}}^\#$ has been restored from the degraded image \mathbf{y} . From this image, we can for instance reduce the blur by applying other filters \mathbf{g}_r and \mathbf{h}_r with smaller spread and visualize the resulting image \mathbf{b}_r , passing the solution $\hat{\mathbf{x}}$ through this new blur operator \mathbf{A}_r ; that is, $\hat{\mathbf{b}}_r = \mathbf{A}_r \hat{\mathbf{x}}$.

5. Recovering the line parameters by a Prony-like method. In this section, we present the method that underlies the second step of this work (see [Figure 3](#)), namely the estimation of the line parameters, which is related to the spectral estimation field. We now focus on estimating the parameters $(\theta_k, \alpha_k, \eta_k)$, which characterize the K lines, from the solution of the minimization problem $\tilde{\mathbf{x}}$ [\(27\)](#), extended by Hermitian symmetry to $m = -M, \dots, -1$ beforehand. This requires the use of a classical spectral estimation method [\[78, 79\]](#).

5.1. Sketch of the 1-D Prony TLS method. The recovering procedure hereafter, based on [\[71\]](#), is an extended method of the famous Prony method [\[68\]](#). Let us sketch this method, which is based on an annihilating property [\[5\]](#). Let $\mathbf{z} = (z_0, \dots, z_{N-1})$ be a complex vector, whose components are:

$$(68) \quad z_i = \sum_{k=1}^K d_k (e^{j2\pi f_k})^i, \quad \forall i = 0, \dots, N-1,$$

with $d_k \in \mathbb{C}$, $f_k \in [-1/2, 1/2)$ the parameters to retrieve and $N \geq 2K + 1$.

Let $\zeta_k = e^{j2\pi f_k}$, we introduce the annihilating polynomial $H(\zeta) = \prod_{l=1}^K (\zeta - \zeta_l) = \sum_{l=0}^K h_l \zeta^{K-l}$ with $h_0 = 1$. Then, we can note that for all $r = K, \dots, N-1$:

$$(69) \quad \sum_{l=0}^K h_l z_{r-l} = \sum_{l=0}^K h_l \left(\sum_{k=1}^K d_k \zeta_k^{r-l} \right) = \sum_{k=1}^K d_k \zeta_k^{r-K} \underbrace{\left(\sum_{l=0}^K h_l \zeta_k^{K-l} \right)}_{H(\zeta_k)=0} = 0.$$

By rearranging the elements z_i in a Toeplitz matrix $\mathbf{P}_K(\mathbf{z})$ of size $(N - K) \times (K + 1)$ and rank K as follows:

$$(70) \quad \mathbf{P}_K(\mathbf{z}) = \begin{pmatrix} z_K & \cdots & z_0 \\ \vdots & \ddots & \vdots \\ z_{N-1} & \cdots & z_{N-K-1} \end{pmatrix},$$

(69) can be written with $\mathbf{h} = (h_0, \dots, h_K)$ as:

$$(\mathbf{z} * \mathbf{h})(r) = 0, \quad \forall r = K, \dots, N - 1 \iff \mathbf{P}_K \mathbf{h} = \mathbf{0}.$$

Consequently, the method consists in finding a right singular vector $\mathbf{h} = (h_0, \dots, h_K)$ of the matrix $\mathbf{P}_K(\mathbf{z})$ associated to the singular value zero. From the SDV decomposition $\mathbf{P}_K(\mathbf{z}) = \mathbf{V}_1 \mathbf{\Sigma} \mathbf{V}_2^*$, it corresponds to the $(K + 1)$ -th column of \mathbf{V}_1 , that is $\mathbf{h} = \mathbf{V}_1[:, K + 1]$. Thus, the roots of the polynomial $H(\zeta) = \sum_{l=0}^K h_l \zeta^{K-l}$ are the searched complex values $\zeta_k = e^{j2\pi f_k}$ and then $f_k = \arg(\zeta_k)/(2\pi)$.

The Algorithm 3 below described the procedure for estimating the frequencies:

Algorithm 3 PRONY

Input: $\tilde{\mathbf{z}} = (\tilde{z}_0, \dots, \tilde{z}_{N-1}) \in \mathbb{C}^N$ a vector of form (68) possibly corrupted by noise, with $N \geq 2K + 1$ and K the number of frequencies to retrieve.

Output: $\{\tilde{f}_k\}_{k=1}^K$ the estimated frequencies.

- 1: Compute the SVD decomposition $\tilde{\mathbf{V}}_1 \tilde{\mathbf{\Sigma}} \tilde{\mathbf{V}}_2^*$ of the matrix $\mathbf{P}_K(\tilde{\mathbf{z}})$ (70)
 - 2: Extract the $(K + 1)$ -th right singular vector $\tilde{\mathbf{h}} = (\tilde{h}_0, \dots, \tilde{h}_K)^T = \tilde{\mathbf{V}}_1[:, K + 1]$.
 - 3: Compute $\tilde{f}_k = \arg(\tilde{\zeta}_k)/(2\pi)$, with $\{\tilde{\zeta}_k\}_{k=1}^K$ the roots of polynomial $\sum_{k=0}^K \tilde{h}_k \zeta^k$.
-

Finally, the complex amplitudes can be retrieved as well, by writing (68) in matrix form $\mathbf{z} = \mathbf{U} \mathbf{d}$, where $\mathbf{d} = (d_1, \dots, d_K)$ and the matrix \mathbf{U} of size $N \times K$ is

$$(71) \quad \mathbf{U} = (\mathbf{a}(f_1) \quad \cdots \quad \mathbf{a}(f_K)) = \begin{pmatrix} 1 & \cdots & 1 \\ e^{-j2\pi f_1} & \cdots & e^{-j2\pi f_K} \\ e^{-j4\pi f_1} & \cdots & e^{-j4\pi f_K} \\ \vdots & \vdots & \vdots \\ e^{-j2\pi(N-1)f_1} & \cdots & e^{-j2\pi(N-1)f_K} \end{pmatrix}$$

and we recover the amplitudes by least-squares approximation:

$$(72) \quad \mathbf{d} = (\mathbf{U}^* \mathbf{U})^{-1} \mathbf{U}^* \mathbf{z}.$$

5.2. Procedure of the line parameters estimation. We start with the angle parameters estimation by applying a Prony-like method onto the rows $\tilde{\mathbf{t}}_m$ of the output solution $\tilde{\mathbf{x}}$, since from (23) and (24) one have theoretically

$$(73) \quad \mathbf{t}_m^\# = \tilde{\mathbf{x}}^\# [m, :] = \sum_{k=1}^K d_{m,k} \mathbf{a}(f_{m,k})^T,$$

whose frequencies to estimate are related to the angles by $f_{m,k} = \frac{\tan \theta_k}{W} m$. The number of lines can be found either by a dedicated method like the *approximate Prony method* (APM) [67] or by evaluating the rank of the dual matrices $\mathbf{T}_{M+1}(\tilde{\mathbf{x}}[:, n_2])$ at the end

of the algorithm, from the decreasing profile of their eigenvalues. The output solution of the algorithm $\tilde{\mathbf{x}}$, whose stopping criteria is met when this one is sufficiently close to the exact solution $\|\tilde{\mathbf{x}} - \hat{\mathbf{x}}^\# \| \ll 1$, then remains some noise. In the noiseless case, the Prony method is always able to recover the frequencies with infinite precision, if the number of samples N is greater than $2K$. But in our case, the estimate $\tilde{f}_{m,k}$ is affected by some uncertainty $\epsilon_{m,k}$; that is, $\tilde{f}_{m,k} = f_{m,k} + \epsilon_{m,k}$, due to the instability of root finding in presence of noise. Then, we propose to estimate the angle parameters by applying K linear regression to the data $\{\tilde{f}_{m,k}\}_{1 \leq m \leq M}$ since

$$\tilde{f}_{m,k} = \frac{\tan \theta_k}{W} m + \epsilon_{m,k} ,$$

which leads to an estimation of the slope $\tan \theta_k$ and then to the angle θ_k . The errors $\epsilon_{m,k}$ committed by evaluating the frequencies $f_{m,k}$ have an amplitude which depends on m . Indeed, for a small m , the frequencies $f_{m,k} = \frac{\tan \theta_k}{W} m$, are close to each other and the Prony method fails to accurately determine the frequencies. Consequently, it is preferable to start the linear regression with the largest values $m \geq m_0$, in order to space the frequencies on the unit circle. We have to make sure that for large values of m , the two extremal frequencies, say $f_{m,1} \leq 0$ and $f_{m,K} \geq 0$, are not close to $-\pi$ and π , respectively, at the same time, as this would violate the separation criteria. Then, the angles $\tilde{\theta}_k$ estimated from the K linear regressions are used to form the matrices $\tilde{\mathbf{U}}_m$ and to obtain the complex estimated amplitudes $\tilde{\mathbf{d}}_m$, by solving least-squares linear systems (72).

REMARK 8. *In the preliminary version of this work [66], we proposed a simplistic method consisting in averaging the estimates*

$$\tilde{\theta}_{m,k} = \arctan(W \tilde{f}_{m,k}/m) = \arctan(W f_{m,k}/m + W \epsilon_{m,k}/m) \approx \theta_{m,k} + W \epsilon_{m,k}/m ,$$

whose error is actually amplified by a factor W/m ; this gives bad results, in particular for a small m . Consequently, the mean $\tilde{\theta}_k = \frac{1}{M} \sum_{m=1}^M \tilde{\theta}_{m,k}$ did not lead to a robust estimation of the angles θ_k .

The previous estimation process of the frequencies $\tilde{f}_{m,k}$ and the angle estimation are possible, as long as the sorting process of the frequencies $\tilde{f}_{m,k}$ is related to the corresponding angles $\theta_1 \leq \dots \leq \theta_k \leq \dots \leq \theta_K$ for all m , which allows us to perform the linear regression with respect to m . It would not be possible to do the same with $\tilde{f}_{n_2,k} = (n_2 \tan \tilde{\theta}_k - \tilde{\eta}_k)/W$ to estimate the offsets η_k , performing the Prony method on the columns, because the affine relation does not preserve the order (one can find n and n' such that $\tilde{f}_{n,k_1} \leq \tilde{f}_{n,k_2}$ and $\tilde{f}_{n',k_1} \geq \tilde{f}_{n',k_2}$).

The frequencies are not uniquely determined, as they belong to an interval of length greater than one $\tilde{f}_{n_2,k} \in [-(H_S + M)/W, (H_S + M)/W]$ and above all we would lose the correspondence between the $\tilde{f}_{n_2,k}$ and the previous estimated angles $\tilde{\theta}_k$, which compromises the estimation of the η_k . Thus, the solution is to perform the Prony method on the estimated vectors $\tilde{\mathbf{e}}_k = \{\tilde{e}_{m,k}\}_{m \geq m_0}^M$, noticing from (24) that

$$d_{m,k}/|d_{m,k}| = \left(e^{-j2\pi\eta_k/W} \right)^m ,$$

which leads to the frequency estimation of a single sampled exponential. This way, the correspondence between the angles $\tilde{\theta}_k$ and the offsets $\tilde{\eta}_k$ is preserved.

Finally, regarding the amplitudes c_k , taking the modulus of the $\tilde{d}_{m,k}$'s leads to inaccurate estimates, since they are computed from the solution $\tilde{\mathbf{x}}$ and their amplitudes have been shrunk, due to the choice of a parameter $c < c^\sharp$ to remove noise. Like often with variational methods, this yields a bias towards zero and yields images with a loss of contrast; that is, the line intensities are globally smaller. That is why we propose to perform a *refitting* step, to recover the amplitudes in a better way. Given the estimated 2-D atoms $\frac{1}{\cos \tilde{\theta}_k} \mathbf{a}_{2D}(\tilde{\theta}_k, \tilde{\eta}_k)$, we evaluate the amplitudes $\tilde{\alpha}_k$ by applying least-squares estimation to the noisy data $\hat{\mathbf{y}}$:

$$(74) \quad (\tilde{\alpha}_1, \dots, \tilde{\alpha}_K) = \arg \min_{\alpha_1, \dots, \alpha_K} \left\| \sum_{k=1}^K \alpha_k \tilde{\mathbf{A}}_k - \hat{\mathbf{y}} \right\|^2, \quad \tilde{\mathbf{A}}_k = \frac{1}{\cos \tilde{\theta}_k} \mathbf{A} \mathbf{a}_{2D}(\tilde{\theta}_k, \tilde{\eta}_k).$$

The procedure is summarized [Algorithm 4](#) and is illustrated in [Figure 5](#):

Algorithm 4 Extraction of line parameters

Input: $\tilde{\mathbf{x}}$ solution of the optimization problem (27) under constraints (28)

Output: $\{\tilde{\theta}_k\}_{k=1}^K$, $\{\tilde{\eta}_k\}_{k=1}^K$ and $\{\tilde{\alpha}_k\}_{k=1}^K$ the estimated parameters of the lines x^\sharp .

- 1: Estimate K via APM or the rank of the matrices $\mathbf{T}_{M+1}(\tilde{\mathbf{x}}[:, n_2])$.
 - 2: **for** $m = 1$ **to** M **do**
 - 3: Extract the row $\tilde{\mathbf{t}}_m = \tilde{\mathbf{x}}[m, :]$.
 - 4: Compute $\{\tilde{f}_{m,k}\}_{k=1}^K \leftarrow \text{PRONY}(\tilde{\mathbf{t}}_m, K)$ and sort the frequencies with resp. to k .
 - 5: **end for**
 - 6: **for** $k = 1$ **to** K **do**
 - 7: Perform a linear regression on $\{\tilde{f}_{m,k}\}_{m \geq m_0}^M$ to estimate $\tan \tilde{\theta}_k$ and then $\tilde{\theta}_k$.
 - 8: **end for**
 - 9: **for** $m = 1$ **to** M **do**
 - 10: Form the matrix $\tilde{\mathbf{U}}_m = (\mathbf{a}(\tan \tilde{\theta}_1 m/W) \cdots \mathbf{a}(\tan \tilde{\theta}_K m/W))$.
 - 11: Compute the vector $\tilde{\mathbf{d}}_m = (\tilde{d}_{m,1}, \dots, \tilde{d}_{m,K})^\top = (\tilde{\mathbf{U}}_m^* \tilde{\mathbf{U}}_m)^{-1} \tilde{\mathbf{U}}_m^* \tilde{\mathbf{t}}_m$.
 - 12: Compute the normalized values $\tilde{e}_{m,k} = \tilde{d}_{m,k} / |\tilde{d}_{m,k}|$.
 - 13: **end for**
 - 14: **for** $k = 1$ **to** K **do**
 - 15: Form the vector $\tilde{\mathbf{e}}_k = (\tilde{e}_{m,k})_{m \geq m_0}^M$.
 - 16: Compute $\tilde{\eta}_k \leftarrow W * \text{PRONY}(\tilde{\mathbf{e}}_k, 1)$.
 - 17: **end for**
 - 18: Evaluate the amplitudes $\{\tilde{\alpha}_k\}_{k=1}^K$ by solving the least-squares problem (74).
-

6. Other related works and further comments. Below we discuss other approaches to estimate line parameters and discuss their effectiveness:

- There exist sophisticated methods, called *debiasing methods* [12, 32, 33], which could be considered to recover the line intensities α_k , instead of the least-squares method (74) used here.

- For recovering the offsets η_k , we also could apply the Prony method on the middle line ($n_2 = 0$) of the image $\tilde{\mathbf{x}}$, since from (22) we have

$$(75) \quad \mathbf{t}_0^\sharp[m] = \sum_{k=1}^K c_k e^{j2\pi\eta_k m/W}$$

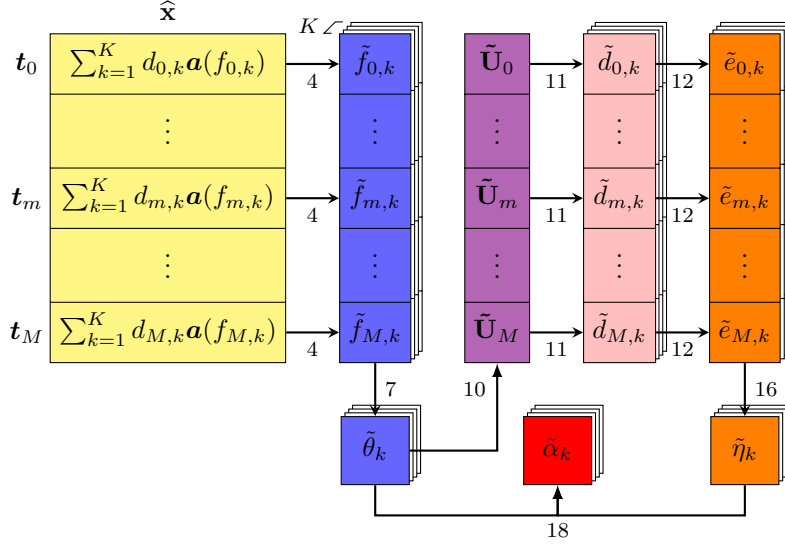


FIG. 5. Parameters extraction procedure illustrating the Algorithm 4.

and since the arguments of these exponentials are uniquely determined. Provided that the latter frequencies are well separated, it would give a better estimation of the offsets $\tilde{\eta}_k$, but the correspondence with the angles $\tilde{\theta}_k$ would be lost, as mentioned previously. To reconnect them, one could rely on the work of [67] (or equivalent [64]), noticing that

$$\hat{\mathbf{x}}^\sharp[m, n_2] = \sum_{k=1}^K \frac{\alpha_k}{\cos \theta_k} e^{j2\pi(n_2 \tan \theta_k - \eta_k) \frac{m}{W}} = \sum_{k=1}^K \frac{\alpha_k}{\cos \theta_k} e^{j2\pi m \left\langle \begin{pmatrix} n_2 \\ -1 \end{pmatrix} \begin{pmatrix} \tan \theta_k / W \\ \eta_k / W \end{pmatrix} \right\rangle}$$

are the samples (indexed by m) of the Fourier transform of $\sum_{k=1}^K \frac{\alpha_k}{\cos \theta_k} \delta(\tan \theta_k / W, \eta_k / W)$ (denoted by h_0) along radial lines indexed by n_2 , that is $h_0(mn_2, m) = \hat{\mathbf{x}}^\sharp[m, n_2]$. One could exploit this information to couple the right frequencies $(\tan \theta_k / W, \eta_k / W)_k$ among all the possibilities $\{(\tan \theta_{k_1} / W, \eta_{k_2} / W) : k_1, k_2 = 1, \dots, K\}$, using the estimated frequencies along the radial lines $h_0(mn_2, m)$; that is, using the Prony method along columns of $\tilde{\mathbf{x}}$.

This procedure, called *sparse approximate Prony method* (SAPM), requires the use of samples along lines $h_0(n, 0)$, $h_0(0, n)$, $h_0(n, \alpha n + \beta)$ for $n \in \mathbb{Z}_N = [-N, N] \cap \mathbb{Z}$ and $\alpha, \beta \in \mathbb{Z}$, conveniently chosen. Therefore, it first requires to separately estimate the frequencies along the horizontal axis $h_0(n, 0)$ and the vertical axis $h_0(0, n)$, before coupling them through the lines $h_0(n, \alpha n + \beta)$. We get the frequencies η_k / W along one axis, since from (75) we have $h_0(0, m) = \mathbf{l}_0^\sharp[m] = \hat{\mathbf{x}}^\sharp[m, 0]$. However, we do not directly have access to $h_0(n, 0)$ from $\hat{\mathbf{x}}^\sharp$ for estimating the $\tan \theta_k / W$. We admittedly have these frequencies appearing on the first row $\hat{\mathbf{x}}^\sharp[1, :]$, but they are clustered and badly separated on this one. Consequently, it is better to use our procedure for estimating the angles, which exploits the other rows for larger index m , in order to space the frequencies before applying the Prony method; this is reminiscent of the strategies of *decimation* developed in [29, 11, 2]. Moreover, our proposed method to estimate the offsets η_k deals with the rows and columns jointly and automatically preserves the correspondence, without having to restore it *a posteriori*. This is less arbitrary than

performing the Prony method on \mathbf{l}_0^\sharp (75), which requires well-separated frequencies on this column; this is a meaningless condition, since the lines can intersect the x -axis.

- The authors of [36] proposed a convex approach to the recovery of a superposition of point sources from samples of its Fourier transform along radial lines. They also emphasize the equivalence between working with the radial Fourier coefficients and working with its Radon projections, according to the Fourier slice theorem. An important question is then how many lines (or projections) and samples are needed to obtain an exact reconstruction. The authors of [74] show that K points can theoretically be determined by the projection onto $K + 1$ distinct lines through the origin. In [64], it was conjectured that under certain additional assumptions, it is possible to choose only four lines passing through the origin to guarantee a unique reconstruction, but this conjecture has been shown to be false in [34]. These authors cast the problem into a nonconvex optimization problem for which there are guarantees for recovering the points, but this constrained optimization problem is NP-hard. Finally, the authors of [36] give a total variation minimization formulation of the problem. Like in [64], where the radial line of angle θ is chosen to maximize the minimal distance between two projections onto it, the authors show the crucial role of the minimal separation distance $\nu_{\min}(\Delta_\theta)$ between the projected positions $\Delta_\theta = \{\langle \theta, x_k \rangle\}_{k=1}^K$. Thus, they proved that for a set of radial measurements $\Theta \in \mathbb{S}^1$ for which the global minimal separation distance is $\nu_{\min} = \inf_{\theta \in \Theta} \nu_{\min}(\Delta_\theta)$, total variation minimization has a unique solution achieved by their algorithm from 3 different radial lines of Θ and a sampling set \mathbb{Z}_N with $N \geq 2/\nu_{\min}$. These results provide strong guarantees, but the framework is quite different from ours, especially because the unit sampling step is constant on every radial line, while in our case the samples are taken on points (mn_2, m) . Moreover, the frequencies along these lines can easily collide, which contradicts the separation assumption on the projected positions. Finally, a common point between our two-step super-resolution recovery of lines and the references cited above is that all these methods come down to 1-D settings, based on a sufficient separation property.

- A final approach might be to apply the Radon transform to the noisy blurred lines $\mathbf{y} = \mathbf{b}^\sharp + \boldsymbol{\epsilon}$, which somehow transform blurred lines to blurred peaks, representing in Figure 1b by 2-D point sources in the Radon domain. Then, we come down to a classical 2-D spikes super-resolution problem, with on the one hand a noise amount which is somehow reduced [58] and on the other hand a minimal separation $\inf_{k \neq l} \text{dist}(\theta_k, \theta_l)$, which is better than the one we have with the frequencies $\tan \theta_k/W$ of $\tilde{\mathbf{x}}$. However, the main drawback is that the Radon transform is performed on the data image \mathbf{y} of blurred and noisy lines, which is finite and discrete, so that it produces some artifacts like “Butterfly patterns” (see Figure 1) [30, 48, 49], due to the finite length of lines. Moreover, they induce discrete other approximations due to the projections on a grid [87]. These problems are avoided with our model and assumptions, which yields an off-the-grid super-resolution estimation of the parameters and outperforms the naive approaches using the Radon (or Hough) transform.

7. Experimental results. The reconstruction procedure was implemented in MATLABTM code, available on the webpage of the first author. We consider an image of size $W = H = 65$, containing three lines of parameters $(\theta_1, \eta_1, \alpha_1) = (-\pi/5, 0, 255)$, $(\theta_2, \eta_2, \alpha_2) = (\pi/16, -15, 255)$ and $(\theta_3, \eta_3, \alpha_3) = (\pi/6, 10, 255)$. We consider the normalized filter \mathbf{h} approximating a Gaussian function of standard deviation κ ; that is, $\varphi_2 : t \mapsto (2\pi\kappa^2)^{-1/2} \exp(-t^2/(2\kappa^2))$, on the compact set $[-S, S]$ with $S = \lceil 4\kappa \rceil - 1$;

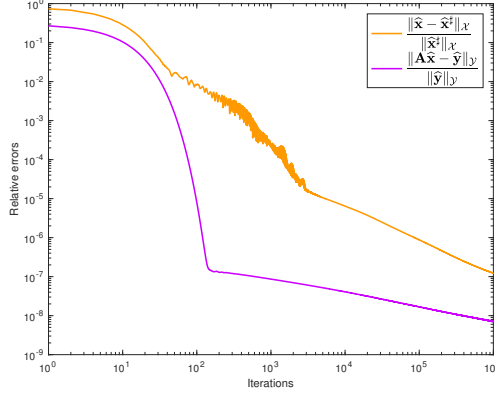


FIG. 6. Log-log plot of the relative errors $\frac{\|\tilde{\mathbf{x}} - \tilde{\mathbf{x}}^\sharp\|_X}{\|\tilde{\mathbf{x}}^\sharp\|_X}$ and $\frac{\|\mathbf{A}\tilde{\mathbf{x}} - \tilde{\mathbf{y}}\|_Y}{\|\tilde{\mathbf{y}}\|_Y}$ for the first experiment.

and the normalized filter $\mathbf{g} = (\mathbf{0}_{M-S}, \mathbf{h}, \mathbf{0}_{M-S})$, whose DFT is an interpolation of $\hat{\mathbf{h}}$, which approaches the continuous Fourier transform $\hat{\varphi}_2 : \nu \mapsto \exp(-2\pi^2 \kappa^2 \nu^2)$. Then, $\|\hat{\mathbf{g}}\|_\infty = \|\hat{\mathbf{h}}\|_\infty = 1$. We use the Algorithm 2 for solving the optimization problem (44) and recovering the lines in the following experiments:

- The first experiment consists in the reconstruction of the lines from $\tilde{\mathbf{x}}$ in absence of noise, (1) by applying the operator \mathbf{A} on this solution, possibly with other kernels \mathbf{g}_r and \mathbf{h}_r and then taking the 1-D inverse Fourier transform; and (2) by applying the Prony method to recover the parameters of the lines, in the aim to display the lines by vectorial drawing. We run the algorithm for 10^6 iterations. Results of relative errors for the solution $\tilde{\mathbf{x}}$ and the estimated parameters are given in Figure 6 and Table 1, where $\Delta_{\theta_i}/\theta_i = |\theta_i - \hat{\theta}_i|/|\theta_i|$, $\Delta_{\alpha_i}/\alpha_i = |\alpha_i - \hat{\alpha}_i|/|\alpha_i|$ and $\Delta_{\eta_i} = |\eta_i - \hat{\eta}_i|$. Although the algorithm is slow to achieve high accuracy, convergence is guaranteed and we observe empirically perfect reconstruction of x^\sharp , when the lines are not too close to each other.

- The purpose of the second experiment is to show the robustness of the method in presence of a strong noise level (Figure 7a). With $c = c^\sharp/3$ and only $2 \cdot 10^3$ iterations, we are able to completely remove the noise and to estimate the line parameters with an error of 10^{-2} .

- Finally, the third experiment, with 10^5 iterations, illustrates the efficiency of the method even in presence of a large blur (Figure 7b), yielding an error of 10^{-4} . For both experiments, the estimated images corresponding to step (1) and (2) are visually identical and are displayed in Figure 7c and Figure 7d.

We emphasize that our algorithm has a much higher accuracy than what is achieved by peak detection after the Hough or Radon transform. These methods are relevant for giving a coarse estimation of line parameters. They are robust to strong noise, but completely fail with a strong blur, which prevents peak detection (see Figure 8). Notice that even by decreasing the discretization steps of the process, we rapidly reach a plateau, as illustrated by Figure 9. This method is limited in accuracy by the pixel grid. By contrast, our super-resolution method makes it possible to achieve infinite precision for the line parameters.

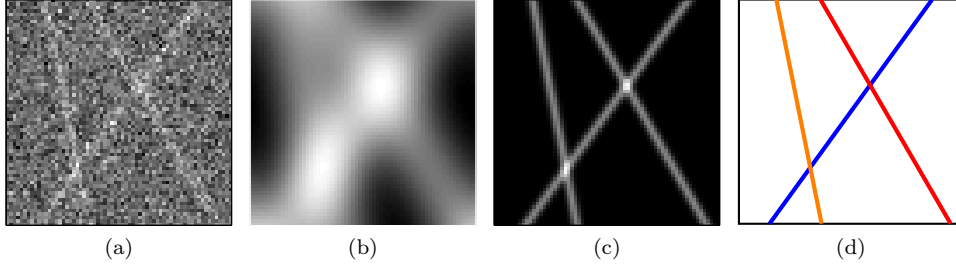


FIG. 7. (a) An image $\mathbf{y} = \mathbf{b}^\sharp + \epsilon$ of three lines x^\sharp , blurred by a Gaussian kernel ϕ with spread $\kappa = 1$ and corrupted by a strong noise $\epsilon \sim \mathcal{N}(0, \zeta)$ with $\zeta = 200$, for the second experiment, (b) with a strong blur ($\kappa = 8$) and no noise ($\zeta = 0$) for the third experiment, (c) the denoised image $\tilde{\mathbf{b}}$, (d) a vectorial drawing of the estimated lines of x^\sharp by the Prony-like method.

TABLE 1
Errors on line parameters recovered by the proposed method.

	Experiment 1	Experiment 2	Experiment 3
Δ_θ/θ	$(10^{-7}, 3.10^{-6}, 7.10^{-7})$	$(10^{-2}, 6.10^{-2}, 9.10^{-2})$	$(6.10^{-7}, 9.10^{-5}, 8.10^{-6})$
Δ_α/α	$(10^{-7}, 10^{-7}, 10^{-7})$	$(10^{-2}, 9.10^{-2}, 2.10^{-1})$	$(4.10^{-5}, 2.10^{-5}, 2.10^{-5})$
Δ_η	$(4.10^{-6}, 7.10^{-6}, 7.10^{-6})$	$(5.10^{-2}, 4.10^{-2}, 3.10^{-2})$	$(5.10^{-5}, 10^{-4}, 3.10^{-4})$

7.1. Closing lines. For a reasonable amount of noise ($\zeta = 20$), the algorithm succeeds in separating two close lines $(\theta_1, \eta_1, \alpha_1) = (-0.73, -1, 255)$ and $(\theta_2, \eta_2, \alpha_2) = (-0.75, 1, 255)$ as illustrated in Figure 10. The estimation of the parameters gives $(\tilde{\theta}_1, \tilde{\eta}_1, \tilde{\alpha}_1) = (-0.725, -0.7, 237)$ and $(\tilde{\theta}_2, \tilde{\eta}_2, \tilde{\alpha}_2) = (-0.753, -0.6, 251)$.

7.2. More lines and different amplitudes. A more complicated example is depicted in Figure 11a, containing seven well-separated lines whose parameters are enumerated in Table 2, corrupted by some noise with variance $\zeta = 20$. We run the algorithm with $c = 0.8c^\sharp$, $\tau = 1$, $\sigma = (\tau(M + H_S + 2))^{-1}$ and after only 2.10^3 iterations, we are able to denoise the image as illustrated in Figure 11b and to estimate the line parameters, with the proposed Prony procedure, as illustrated in Figure 11d, with an error of 10^{-2} , as reported in Table 3.

7.3. General case. We consider an image $\mathbf{y} = \mathbf{b}^\sharp + \epsilon$ composed as four noisy blurred lines ($\kappa = 1$ and $\zeta = 20$) whose two of them are rather vertical (i.e with angle in $(-\pi/4, \pi/4]$), while the two other ones are rather horizontal. The extended algorithm presented in subsection 4.3, provides after $n = 6 \cdot 10^4$ iterations the denoised images depicted in Figure 12. It acts as an angles selector according to the horizontality and verticality: the rather vertical lines are gathered in the reconstructed image $\tilde{\mathbf{b}}_1$ (see Figure 12b) and the rather horizontal lines are gathered in the reconstructed image $\tilde{\mathbf{b}}_2$ (see Figure 12c), so that $\mathbf{b}^\sharp \approx \tilde{\mathbf{b}}_1 + \tilde{\mathbf{b}}_2$, which validates the procedure.

7.4. Inpainting. We consider occluded lines with a mask \mathbf{M} as described in the subsection 4.4. In Figure 13, we occlude a portion of the blurred line ($\kappa = 1$) by applying a square mask \mathbf{M} in the middle. We run the optimization algorithm and visualize the Figure 13a to Figure 13d, that is the evolution of the reconstruction after $n = 2.10^3$, $n = 10^4$ and $n = 10^6$ iterations. We can see that the method is able

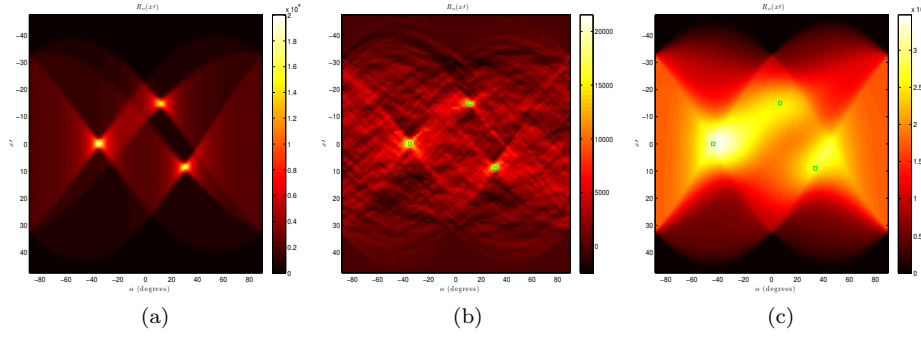


FIG. 8. The Radon transform of the image \mathbf{y} for Experiments 1, 2 and 3. The true parameters of the lines are in green.

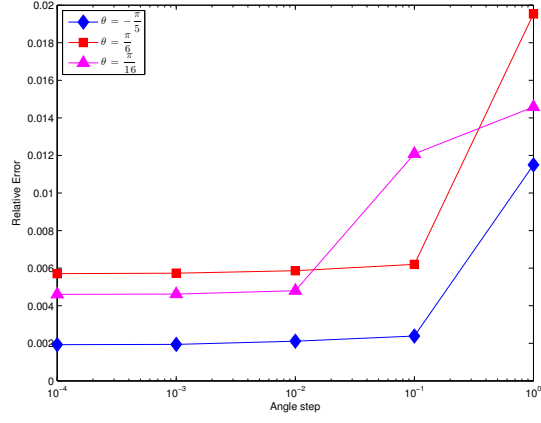


FIG. 9. Change of accuracy with respect to the angle step of the MATLABTM Radon transform.

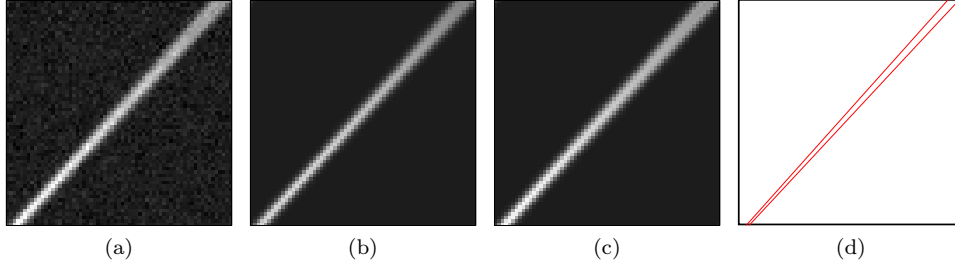


FIG. 10. (a) An image $\mathbf{y} = \mathbf{b}^\sharp + \epsilon$ of two closed lines x^\sharp , blurred by a Gaussian kernel ϕ with spread $\kappa = 1$ and corrupted by noise $\epsilon \sim \mathcal{N}(0, 20)$ (b) the denoised image $\tilde{\mathbf{b}}$, (c) the ground truth image \mathbf{b}^\sharp (d) a vectorial drawing of the estimated lines of x^\sharp by the Prony-like method.

TABLE 2
Angles, offsets and amplitudes of the seven lines.

θ_k	-0.75	-0.5	-0.25	10^{-3}	0.3	0.55	0.75
η_k	15	25	2	7	-20	-5	-10
α_k	60	80	255	100	180	120	240

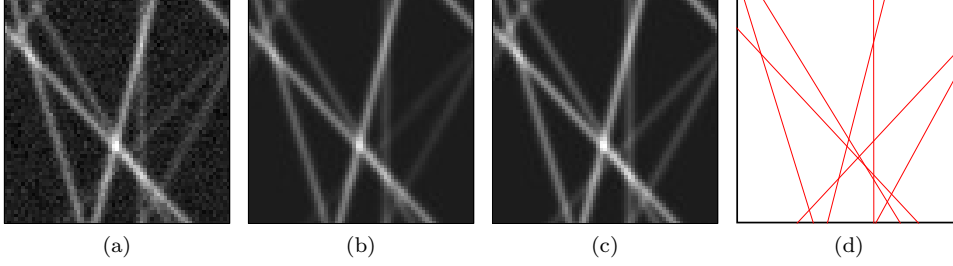


FIG. 11. (a) An image $\mathbf{y} = \mathbf{b}^\sharp + \epsilon$ of seven well-separated lines x^\sharp , blurred by a Gaussian kernel ϕ with spread $\kappa = 1$ and corrupted by noise $\epsilon \sim \mathcal{N}(0, 20)$ (b) the denoised image $\tilde{\mathbf{b}}$, (c) the ground truth image \mathbf{b}^\sharp (d) a vectorial drawing of the estimated lines of x^\sharp by the Prony-like method.

TABLE 3
Errors on line parameters recovered by the proposed method.

$\Delta\theta$	1.10^{-2}	2.10^{-2}	1.10^{-3}	2.10^{-3}	5.10^{-3}	5.10^{-3}	1.10^{-3}
$\Delta\eta$	5.10^{-1}	7.10^{-2}	4.10^{-2}	1.10^{-1}	1.10^{-2}	2.10^{-2}	1.10^{-2}
$\Delta\alpha/\alpha$	4.10^{-2}	5.10^{-2}	5.10^{-3}	4.10^{-2}	6.10^{-3}	1.10^{-2}	4.10^{-3}

to reconstruct the part of line occluded thanks to the information available outside the mask and to the optimization constraints related to the line structure. In the same way, Figure 13e to Figure 13h enable to visualize the “diffusion process” of the information within the occluded part in the Fourier domain. Finally, the experiment is carried out for an image containing three occluded lines, either by a bigger mask in Figure 14a, or by a mask whose binary inputs were drawn randomly Figure 14b. In both cases the lines are successfully reconstructed as in Figure 14c.

8. Conclusion. We proposed a new variational formulation for the problem of recovering lines in degraded images, using the framework of atomic norm minimization. A primal-dual splitting algorithm has been used to solve the convex optimization problem. We applied it successfully to several image restoration problems, recovering lines parameters by the Prony method and we showed the robustness of the method to strong blur and strong noise level. We insist on the novelty of our approach, which is to estimate lines with parameters (angle, offset, amplitude) living in a continuum, with perfect reconstruction in absence of noise, without being limited by the discrete nature of the image, nor its finite size. This work can be viewed as a proof of concept for super-resolution line detection and invites us to revisit the Hough transform in a continuous way. Many theoretical questions remain open, like the study of the separation conditions under which perfect reconstruction can be guaranteed. The robustness of the method needs to be theoretically studied and would requires a statistical analysis ; this is left for future work. From a practical point of view, parallel computing would be welcome to speed up the proposed algorithm. We should also investigate the possibility of relaxing the periodicity and bandlimitedness assumptions, possibly by solving a convex feasibility program without any regularizers [39, 38]. At the time of finalizing this paper, we became aware of the two other recent papers [7] and [40]; we leave for future work the study of their relationship to the setting considered here. In subsequent work, we plan to apply the proposed approach to biomedical images containing curved structures, like tubulins, by operating on small overlapping patches.

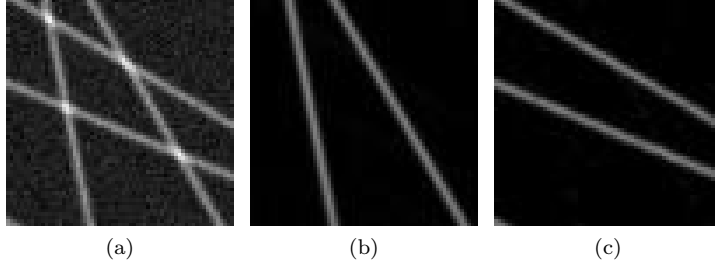


FIG. 12. (a) An image $\mathbf{y} = \mathbf{b}^\sharp + \epsilon$ of four well-separated lines x^\sharp , blurred by a Gaussian kernel ϕ with spread $\kappa = 1$ and corrupted by noise $\epsilon \sim \mathcal{N}(0, 20)$, two of them being rather vertical (angle in $(-\pi/4, \pi/4]$) and the two other ones rather horizontal, (b) the reconstructed image $\tilde{\mathbf{b}}_1$, (c) the reconstructed image $\tilde{\mathbf{b}}_2$.

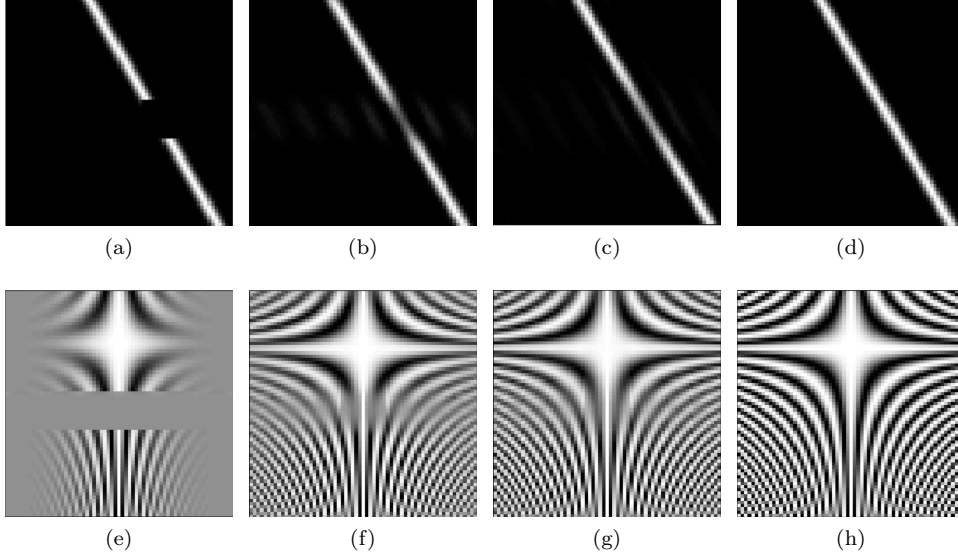


FIG. 13. (a) An image $\mathbf{y} = \mathbf{b}^\sharp + \epsilon$ of one line x^\sharp , blurred by a Gaussian kernel ϕ with spread $\kappa = 1$, no noise and a small square mask \mathbf{M} , (b) reconstruction of $\tilde{\mathbf{b}}$ by inpainting after $2 \cdot 10^3$ iterations, (c) after 10^4 iterations, (d) at convergence ; (e)-(f)-(g)-(h) the corresponding reconstructions of $\hat{\mathbf{x}}$ in the Fourier domain over iterations.

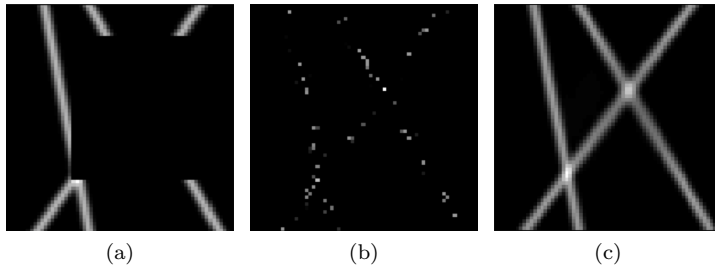


FIG. 14. (a) An image $\mathbf{y} = \mathbf{b}^\sharp + \epsilon$ of three line x^\sharp , blurred by a Gaussian kernel ϕ with spread $\kappa = 1$, no noise and a large square mask \mathbf{M} , (b) with a random mask and (c) their reconstruction at convergence.

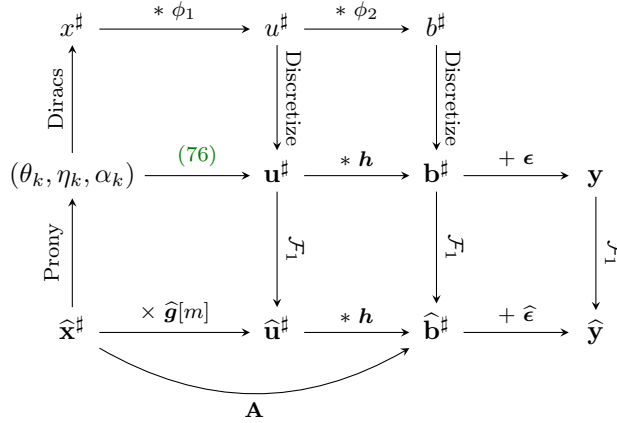


FIG. 15. Relations between the variables.

Appendix A. Characterization of the sampling process.

Formally, we can write $x^\# * \phi = (x^\# * \phi_1) * \phi_2$ with $\phi_1(t_1, t_2) = \varphi_1(t_1)\delta(t_2)$ and $\phi_2(t_1, t_2) = \delta(t_1)\varphi_2(t_2)$, where φ_1 and φ_2 are L^1 functions. So, after the first horizontal convolution, using the fact that $\delta(at) = \delta(t)/|a|$ for any $a \neq 0$, we obtain the function:

$$(76) \quad u^\# = x^\# * \phi_1 : (t_1, t_2) \mapsto \sum_{k=1}^K \frac{\alpha_k}{\cos \theta_k} \varphi_1(t_1 - \eta_k + t_2 \tan \theta_k).$$

We can show that, after the second vertical convolution, we get the function $b^\#$ obtained in (6). Figure 15 explains our notations in more details and illustrates the relation between all continuous and discrete variables. In the following we consider a weaker version of Assumption (ii) about φ_2 :

Assumption (ii') $\varphi_2 \in L^1(\mathbb{R})$ is such that $\int_{\mathbb{R}} \varphi_2(t_2) dt_2 = 1$. Denoting $\text{sinc}(t_2) = \sin(\pi t_2)/(\pi t_2)$ the sinus cardinal with $\text{sinc}(0) = 1$, we make the assumption that the discrete filter

$$(77) \quad (\mathbf{h}[n] = (\varphi_2 * \text{sinc})(n))_{n \in \mathbb{Z}},$$

has compact support of length $2S + 1$, for some $S \in \mathbb{N}$; that is,

$$\mathbf{h}[n] = 0, \quad \text{if } |n| \geq S + 1.$$

REMARK 9. Note that if φ_2 is bandlimited, we simply have $\mathbf{h}[n] = \varphi_2(n)$, which reverts back to Assumption (ii).

Let us deduce from these assumptions some other properties satisfied by φ_1 and φ_2 and their associated discrete filters \mathbf{g} and \mathbf{h} . First, we have the following proposition:

PROPOSITION 11 (Nyquist–Whittaker–Shannon). *The function φ_1 , which is periodic and bandlimited, is determined by W degrees of freedom only. That is, with the*

coefficients $\mathbf{g}[n] = \varphi_1(n)$, $n = 0, \dots, W-1$, the function φ_1 is a linear combination of shifted Dirichlet kernels $D_M(t) = \sum_{m=-M}^M e^{jmt}$:

$$(78) \quad \varphi_1(t) = \frac{1}{W} \sum_{n=0}^{W-1} \mathbf{g}[n] D_M\left(\frac{2\pi(t-nW)}{W}\right), \quad \forall t \in \mathbb{R}.$$

Proof. Let us start with the following classical theorem; in our notations:

THEOREM 12 (Nyquist Whittaker-Shannon). *Let f be a T -periodic function and $c_m(f) = 0$ for $|m| \geq M+1$. Then f can be reconstructed from the regular sampling $\{f(ka), k = 0, 1, \dots, 2M\}$, where $a = \frac{T}{2M+1}$ is the sampling rate, in this way:*

$$(79) \quad f(x) = \frac{1}{2M+1} \sum_{k=0}^{2M} f(ka) D_M\left(\frac{2\pi}{T}(x-ka)\right),$$

where D_M is the Dirichlet Kernel:

$$(80) \quad D_M(x) = \sum_{m=-M}^M e^{jmx} = \frac{\sin\left((M+\frac{1}{2})x\right)}{\sin\frac{x}{2}}.$$

Applying to φ_1 which is W -periodic and $c_m(\varphi_1) = 0$ for $|m| \geq M+1$ with $2M+1 = W$, then $a = 1$ and the Nyquist theorem leads to:

$$(81) \quad \varphi_1(t) = \sum_{n=0}^{W-1} \varphi_1(n) \frac{\sin \pi(t-n)}{W \sin\left(\frac{\pi(t-n)}{W}\right)}.$$

Consequently, we give this explicit formula $\mathbf{g}[n] = \varphi_1(n)$ and $\widehat{\mathbf{g}}[n] = c_{n-M}(\varphi_1)$ for $n = 0, 1, \dots, 2M$. By the simple change of variable $n \leftarrow n - M$, one can also obtain:

$$(82) \quad \sum_{n=-M}^M \varphi_1(n) \frac{\sin \pi(t-n)}{W \sin\left(\frac{\pi(t-n)}{W}\right)}. \quad \square$$

Then,

$$\frac{1}{W} \int_0^W \varphi_1(t) dt = \frac{1}{W} \sum_{n=-M}^M \mathbf{g}[n] = 1,$$

that is, the filter \mathbf{g} is normalized as well. Moreover

$$\int_0^W \varphi_1(t)^2 dt = \sum_{n=-M}^M \mathbf{g}[n]^2 = \sum_{n=-M}^M \varphi_1(n)^2,$$

and by Parseval relation:

$$\sum_{m \in \mathbb{Z}} |c_m(\varphi_1)|^2 = \sum_{m=-M}^M |\widehat{\mathbf{g}}[m]|^2 = \frac{1}{W} \int_0^W \varphi_1(t)^2 dt.$$

Now, we describe the sampling process leading from continuous to discrete formulation, based on the following proposition:

PROPOSITION 13. *It is equivalent to perform the vertical convolution of $u^\sharp = x^\sharp * \phi_1$ with φ_2 , with $\varphi_2 * \text{sinc}$, or with the Dirac comb $\gamma : t_2 \mapsto \sum_{n=-S}^S \mathbf{h}[n] \delta(t_2 - n)$, where $\mathbf{h}[n] = (\varphi_2 * \text{sinc})(n)$.*

Proof. Due to Hypothesis 1, the assumption $\theta_k \in (-\pi/4, \pi/4]$ yields $|\tan \theta_k| \leq 1$ for every $k = 1, \dots, K$. So, the function u^\sharp given in (76), as a function of t_2 at fixed t_1 , is bandlimited: for every $t_1 \in [0, W)$, the Fourier transform $\mathcal{F}_2 u^\sharp : \omega_2 \mapsto \int_{\mathbb{R}} u^\sharp(t_1, t_2) e^{-j2\pi\omega_2 t_2} dt_2$, which is a distribution (sum of K Dirac combs), is zero for every $|\omega_2| \geq 1/2$. Indeed, we have:

$$[\mathcal{F}_2 u^\sharp](\omega_2) = \sum_{k=1}^K \frac{\alpha_k}{\sin \theta_k} \hat{\varphi}_1 \left(\frac{\omega_2}{\tan \theta_k} \right) \exp \left(j2\pi\omega_2 \frac{t_1 - \eta_k}{\tan \theta_k} \right).$$

Since $|\tan \theta_k| \leq 1$, we have $|\omega_2 / \tan \theta_k| \geq |\omega_2|$. The support of $\hat{\varphi}_1$ is included in $[-1/2, 1/2]$ ($c_m(\varphi_1) = 0$ for $|m| \geq M+1$ and $M/W < 1/2$), as well as the support of $\mathcal{F}_2 u^\sharp$ which is necessarily included in the support of $\hat{\varphi}_1$. Then, we have the equivalence $\mathcal{F}_2 u^\sharp = \mathcal{F}_2 u^\sharp \cdot \mathbb{1}_{[-1/2, 1/2]} \Leftrightarrow u^\sharp = u^\sharp * \text{sinc}$ and furthermore $u^\sharp * \varphi_2 = u^\sharp * (\varphi_2 * \text{sinc})$. In the Fourier domain, the function $h = \varphi_2 * \text{sinc}$ is bandlimited, so $[\mathcal{F}_2 u^\sharp] \hat{h} = [\mathcal{F}_2 u^\sharp] \hat{h}_{\text{per}}$ where \hat{h}_{per} corresponds to the periodization of the spectrum of \hat{h} with period 1, which amounts to saying that

$$u^\sharp * h = u^\sharp * \left(\sum_n h[n] \delta(\cdot - n) \right).$$

□

REMARK 10. *Assumption (ii') implies that the filter $(\mathbf{h}[n])_n$ should have compact support, but we can note that the function $h = \varphi_2 * \text{sinc}$ does not have compact support, since it is bandlimited. This means that the continuous function h has to vanish at integer points $t = n$ for $|n| > S$. Given such a compact filter $(\mathbf{h}[n])_{n=-S}^S$, the unique bandlimited function h satisfying these conditions is obtained by the Shannon interpolation formula:*

$$h(t) = \sum_{n=-S}^S \mathbf{h}[n] \text{sinc}(t - n).$$

*By uniqueness, we necessarily have $\varphi_2 * \text{sinc} = h$ and we can notice that there always exists a bandlimited solution φ_2 of this equation, which is simply $\varphi_2 = h$. In practice, we can always approximate a PSF by a bandlimited function h , with $2S+1$ samples $\mathbf{h}[n]$ of this PSF; that is why we argued in Assumption (ii) that the compact support assumption is not restrictive.*

Now, to obtain the discrete image \mathbf{b}^\sharp of (5), let us first define \mathbf{u}^\sharp by sampling $u = x^\sharp * \phi_1$ with unit step:

$$(83) \quad \mathbf{u}^\sharp[n_1, n_2] = (x^\sharp * \phi_1)(n_1, n_2), \quad \forall n_1 = 0, \dots, W-1, \quad n_2 = -S, \dots, H-1+S.$$

With the above assumptions and Proposition 13, we can express \mathbf{b}^\sharp from \mathbf{u}^\sharp using a discrete vertical convolution with the filter \mathbf{h} :

$$(84) \quad \mathbf{b}^\sharp[n_1, n_2] = \sum_{p=-S}^S \mathbf{u}^\sharp[n_1, n_2 - p] \mathbf{h}[p], \quad \forall n_1 = 0, \dots, W-1, \quad n_2 = 0, \dots, H-1.$$

Altogether, we completely and exactly characterized the sampling process, which involves a continuous blur ϕ , using the two discrete and finite filters $(g[n])_{n=0}^{W-1}$ and $(h[n])_{n=-S}^S$.

Appendix B. Proof of Proposition 1. First, we consider the image $\hat{\mathbf{u}}^\#$ obtained by applying the 1-D DFT on every column of $\mathbf{u}^\#$ (see (83)):

$$(85) \quad \hat{\mathbf{u}}^\#[m, n_2] = \frac{1}{W} \sum_{n_1=0}^{W-1} \mathbf{u}^\#[n_1, n_2] e^{-j \frac{2\pi m}{W} n_1},$$

$$\forall m = -M, \dots, M, \quad n_2 = -S, \dots, H-1+S,$$

which are the exact Fourier coefficients of the function $t \mapsto (x^\# * \phi_1)(t, n_2)$, following Assumption (i). Hence, from (76) and $\hat{\mathbf{u}}^\#[m, n_2] = \frac{1}{W} \int_0^W (x^\# * \phi_1)(t, n_2) e^{-j \frac{2\pi m}{W} t} dt$, we obtain:

$$(86) \quad \hat{\mathbf{u}}^\#[m, n_2] = \hat{\mathbf{g}}[m] \hat{\mathbf{x}}^\#[m, n_2], \quad \forall m = -M, \dots, M, \quad n_2 = -S, \dots, H-1+S.$$

Now we apply a 1-D DFT on the first component of the discrete image $\mathbf{b}^\#$ (see (84)), leading to the elements

$$(87) \quad \hat{\mathbf{b}}^\#[m, n_2] = (\hat{\mathbf{u}}^\#[m, :] * \mathbf{h})[n_2], \quad \forall m = -M, \dots, M, \quad n_2 = 0, \dots, H-1.$$

Following (86) and (87), as illustrated in Figure 15, we have

$$\mathbf{A} \hat{\mathbf{x}}^\# = \hat{\mathbf{b}}^\#,$$

which concludes the proof.

Appendix C. Proof of Proposition 3.

Let us introduce

$$(88) \quad \text{SDP}(\mathbf{z}) = \inf_{\mathbf{q} \in \mathbb{C}^N, q_0 \geq 0} \left\{ q_0 : \mathbf{T}'_N(\mathbf{z}, \mathbf{q}) = \begin{pmatrix} \mathbf{T}_N(\mathbf{q}) & \mathbf{z} \\ \mathbf{z}^* & q_0 \end{pmatrix} \succcurlyeq 0 \right\}.$$

We want to prove that $\text{SDP}(\mathbf{z}) = \|\mathbf{z}\|_{\mathcal{A}}$ and that the minimum in (88) is achieved.

- Suppose that $\mathbf{z} = \sum_{k=1}^K c_k \mathbf{a}(f_k, \phi_k)$, with $c_k > 0$.

Let us define $\mathbf{q} = \sum_{k=1}^K c_k \mathbf{a}(f_k)$, with $\mathbf{q} = (q_0, q_1, \dots, q_{N-1})$. Then $q_0 = \sum_{k=1}^K c_k$. For $i = 0, \dots, N-1$, the atoms $\mathbf{a}(f_k)$ have elements $[\mathbf{a}(f_k)]_i = e^{j2\pi f_k i}$, hence

$$\begin{aligned} \mathbf{T}_N(\mathbf{a}(f_k)) &= \begin{pmatrix} 1 & e^{-j2\pi f_k} & \dots & e^{-j2\pi f_k(N-1)} \\ e^{j2\pi f_k} & 1 & \dots & e^{-j2\pi f_k(N-2)} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j2\pi f_k(N-1)} & e^{j2\pi f_k(N-2)} & \dots & 1 \end{pmatrix}, \\ &= \begin{pmatrix} 1 \\ e^{j2\pi f_k} \\ \vdots \\ e^{j2\pi f_k(N-1)} \end{pmatrix} \begin{pmatrix} 1 & e^{-j2\pi f_k} & \dots & e^{-j2\pi f_k(N-1)} \end{pmatrix}, \\ &= \mathbf{a}(f_k) \mathbf{a}(f_k)^*. \end{aligned}$$

We deduce that

$$\begin{aligned}\mathbf{T}_N(\mathbf{q}) &= \sum_{k=1}^K c_k \mathbf{T}(\mathbf{a}(f_k)) , \\ &= \sum_{k=1}^K c_k \mathbf{a}(f_k) \mathbf{a}(f_k)^* , \\ &= \sum_{k=1}^K c_k \mathbf{a}(f_k, \phi_k) \mathbf{a}(f_k, \phi_k)^* .\end{aligned}$$

Therefore, the matrix

$$\begin{pmatrix} \mathbf{T}_N(\mathbf{q}) & \mathbf{z} \\ \mathbf{z}^* & q_0 \end{pmatrix} = \sum_{k=1}^K c_k \begin{pmatrix} \mathbf{a}(f_k, \phi_k) \\ 1 \end{pmatrix} \begin{pmatrix} \mathbf{a}(f_k, \phi_k) \\ 1 \end{pmatrix}^*$$

is positive semidefinite. Given $q_0 = \sum_{k=1}^K c_k$, we get $\text{SDP}(\mathbf{z}) \leq \sum_{k=1}^K c_k$. Since this holds for any decomposition of \mathbf{z} , we conclude that $\text{SDP}(\mathbf{z}) \leq \|\mathbf{z}\|_{\mathcal{A}}$.

- Conversely, let $\mathbf{q} \in \mathbb{C}^N$ be a vector such that $q_0 \geq 0$ and $\begin{pmatrix} \mathbf{T}_N(\mathbf{q}) & \mathbf{z} \\ \mathbf{z}^* & q_0 \end{pmatrix} \succcurlyeq 0$.

In particular we have $\mathbf{T}_N(\mathbf{q}) \succcurlyeq 0$. We denote by r the rank of $\mathbf{T}_N(\mathbf{q})$. [Theorem 2](#) ensures that $\mathbf{T}_N(\mathbf{q}) \succcurlyeq 0$ and is of rank $r \leq N$, if and only if there exists $d_k > 0$ and distinct f_k , such that

$$(89) \quad \mathbf{q} = \sum_{k=1}^r d_k \mathbf{a}(f_k) ,$$

$$(90) \quad q_0 = \sum_{k=1}^r d_k .$$

Let us set $\mathbf{D} = \text{diag}(d_1, \dots, d_r)$ and

$$\mathbf{V} = (\mathbf{a}(f_1) \quad \dots \quad \mathbf{a}(f_r)) = \begin{pmatrix} 1 & 1 & \dots & 1 \\ e^{j2\pi f_1} & e^{j2\pi f_2} & \dots & e^{j2\pi f_r} \\ e^{j2\pi f_1^2} & e^{j2\pi f_2^2} & \dots & e^{j2\pi f_r^2} \\ \vdots & \vdots & \vdots & \vdots \\ e^{j2\pi f_1(N-1)} & e^{j2\pi f_2(N-1)} & \dots & e^{j2\pi f_r(N-1)} \end{pmatrix} .$$

By linearity of the operator \mathbf{T}_N :

$$\begin{aligned}\mathbf{T}_N(\mathbf{q}) &= \sum_{k=1}^r d_k \mathbf{T}_N(\mathbf{a}(f_k)) , \\ &= \sum_{k=1}^r d_k \mathbf{a}(f_k) \mathbf{a}(f_k)^* , \\ &= \mathbf{V} \mathbf{D} \mathbf{V}^* .\end{aligned}$$

Since $\mathbf{T}_N(\mathbf{a}(f_k))$ contains only ones on the diagonal, we have

$$\frac{1}{N} \text{tr}(\mathbf{T}_N(\mathbf{q})) = \sum_{k=1}^r d_k > 0 .$$

Besides, $\frac{1}{N} \text{tr}(\mathbf{T}_N(\mathbf{q})) = q_0$, therefore $q_0 > 0$.

Let be \mathbf{M} a general block matrix $\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{C} \end{pmatrix}$, the Schur complement gives

$$[\mathbf{C} \succ 0 \Rightarrow \mathbf{M} \succ 0] \implies [\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^* \succ 0].$$

We apply this lemma to $\mathbf{M} = \begin{pmatrix} \mathbf{T}_N(\mathbf{q}) & \mathbf{z} \\ \mathbf{z}^* & q_0 \end{pmatrix}$, with $\mathbf{A} = \mathbf{T}_N(\mathbf{q})$, $\mathbf{B} = \mathbf{z}$ and $\mathbf{C} = q_0$. The left term is satisfied by hypothesis, hence

$$\mathbf{T}_N(\mathbf{q}) - q_0^{-1}\mathbf{z}\mathbf{z}^* \succ 0 \iff \mathbf{V}\mathbf{D}\mathbf{V}^* - q_0^{-1}\mathbf{z}\mathbf{z}^* \succ 0.$$

We define the square matrix \mathbf{V}_r by extracting the r first rows and columns of \mathbf{V} , which is a Vandermonde matrix, whose determinant is

$$\det(\mathbf{V}_r) = \prod_{1 \leq k < l \leq r} (\mathbf{a}(f_l) - \mathbf{a}(f_k)).$$

Since we assumed $f_k \neq f_l$, $\forall k \neq l$, \mathbf{V}_r is invertible and $\text{rank}(\mathbf{V}) = r$. Let us define $\mathbf{v} : \mathbb{C}^r \rightarrow \mathbb{C}^N$ and $\mathbf{v}^* : \mathbb{C}^N \rightarrow \mathbb{C}^r$ the linear operators corresponding to matrices \mathbf{V} and \mathbf{V}^* . We have $\text{rank}(\mathbf{v}^*) = \text{rank}(\mathbf{v}) = r$. By the rank-nullity theorem:

$$\dim(\ker \mathbf{v}^*) = N - r.$$

Thus, there exists a vector $\mathbf{p} \in \mathbb{C}^N$ such that $\mathbf{p} \neq \mathbf{0}$ and $\mathbf{V}^*\mathbf{p} = \mathbf{0} \Leftrightarrow \mathbf{p}^*\mathbf{V} = \mathbf{0}$. Consequently,

$$\begin{aligned} \mathbf{p}^*(\mathbf{V}\mathbf{D}\mathbf{V}^* - q_0^{-1}\mathbf{z}\mathbf{z}^*)\mathbf{p} &\geq 0 \Leftrightarrow (\mathbf{p}^*\mathbf{V})\mathbf{D}(\mathbf{V}^*\mathbf{p}) - q_0^{-1}\mathbf{p}^*\mathbf{z}\mathbf{z}^*\mathbf{p} \geq 0, \\ &\Leftrightarrow q_0^{-1}\|\mathbf{p}^*\mathbf{z}\|_2^2 \leq 0, \\ &\Leftrightarrow \|\mathbf{p}^*\mathbf{z}\|_2^2 = 0, \\ &\Leftrightarrow \mathbf{p}^*\mathbf{z} = 0, \\ &\Leftrightarrow \mathbf{p} \perp \mathbf{z}. \end{aligned}$$

Since $\mathbf{p} \in \ker \mathbf{v}^*$, then $\mathbf{z} \in (\ker \mathbf{v}^*)^\perp = \text{Im } \mathbf{v}$, so there exists a vector $\mathbf{w} \in \mathbb{C}^r$ such that $\mathbf{z} = \mathbf{V}\mathbf{w} = \sum_{k=1}^r w_k \mathbf{a}(f_k)$, hence

$$\mathbf{V}\mathbf{D}\mathbf{V}^* - q_0^{-1}\mathbf{V}\mathbf{w}\mathbf{w}^*\mathbf{V}^* \succ 0.$$

Besides, $\text{Im } \mathbf{v}^* \subset \mathbb{C}^r$ and $\dim(\text{Im } \mathbf{v}^*) = \text{rank}(\mathbf{v}^*) = r = \dim(\mathbb{C}^r)$, thus $\text{Im } \mathbf{v}^* = \mathbb{C}^r$ and \mathbf{v}^* is surjective. Consequently, there exists a vector $\mathbf{u} \in \mathbb{C}^N$ such that $\mathbf{V}^*\mathbf{u} = \text{sgn}(\mathbf{w}) = (w_1/|w_1|, \dots, w_r/|w_r|)^\top$ and

$$\begin{aligned} \mathbf{u}^*(\mathbf{V}\mathbf{D}\mathbf{V}^* - q_0^{-1}\mathbf{V}\mathbf{w}\mathbf{w}^*\mathbf{V}^*)\mathbf{u} &\geq 0 \Leftrightarrow (\mathbf{u}^*\mathbf{V})\mathbf{D}(\mathbf{V}^*\mathbf{u}) - q_0^{-1}(\mathbf{u}^*\mathbf{V})\mathbf{w}\mathbf{w}^*(\mathbf{V}^*\mathbf{u}) \geq 0, \\ &\Leftrightarrow \text{sgn}(\mathbf{w})^*\mathbf{D}\text{sgn}(\mathbf{w}) - \frac{1}{q_0}\text{sgn}(\mathbf{w})^*\mathbf{w}\mathbf{w}^*\text{sgn}(\mathbf{w}) \geq 0, \\ &\Leftrightarrow \sum_{k=1}^r d_k \left| \frac{w_k}{|w_k|} \right|^2 - q_0^{-1} \left(\sum_{k=1}^r \frac{w_k^*}{|w_k|} w_k \right)^2 \geq 0, \\ &\Leftrightarrow q_0^2 \geq \left(\sum_{k=1}^r |w_k| \right)^2, \text{ (since } q_0 = \sum_{k=1}^r d_k), \\ &\Leftrightarrow q_0 \geq \sum_{k=1}^r |w_k| \geq \|\mathbf{z}\|_{\mathcal{A}}, \end{aligned}$$

by definition of the atomic norm (19). Taking the infimum leads to $\text{SDP}(\mathbf{z}) \geq \|\mathbf{z}\|_{\mathcal{A}}$.

• Finally, let us show that the infimum of the linear form $\ell : \mathbf{q} \mapsto q_0$ is achieved on the set

$$A(\mathbf{z}) = \left\{ \mathbf{q} \in \mathbb{R}^+ \times \mathbb{C}^{N-1} : \mathbf{T}'_N(\mathbf{z}, \mathbf{q}) = \begin{pmatrix} \mathbf{T}_N(\mathbf{q}) & \mathbf{z} \\ \mathbf{z}^* & q_0 \end{pmatrix} \succcurlyeq 0 \right\} ;$$

that is,

$$(91) \quad \text{SDP}(\mathbf{z}) = \inf_{\mathbf{q} \in A(\mathbf{z})} \ell(\mathbf{q}) = \min_{\mathbf{q} \in A(\mathbf{z})} \ell(\mathbf{q}) .$$

(i) Let us notice that since $\mathbf{q} \in A(\mathbf{z})$ implies $\mathbf{T}_N(\mathbf{q}) \succcurlyeq 0$ then

$$\ell(\mathbf{q}) = q_0 = \frac{1}{N} \text{tr}(\mathbf{T}_N(\mathbf{q})) = \sum_{i=0}^{N-1} \lambda_i \geq 0 ,$$

with λ_i the eigenvalues of $\mathbf{T}_N(\mathbf{q})$ which are positive reals.

(ii) First we show that $A(\mathbf{z})$ is nonempty, since $\mathbf{q} = (\|\mathbf{z}\|_2, 0, \dots, 0)^\top \in A(\mathbf{z})$. Indeed for a fixed vector $\mathbf{z} = (z_0, \dots, z_{N-1}) \in \mathbb{C}^N$, $\mathbf{v} = (v_0, \dots, v_N) \in \mathbb{C}^{N+1}$ and $\mathbf{v}' = (v_0, \dots, v_{N-1}) \in \mathbb{C}^N$ we have for this \mathbf{q} :

$$\begin{aligned} \mathbf{v}^* \begin{pmatrix} \mathbf{T}_N(\mathbf{q}) & \mathbf{z} \\ \mathbf{z}^* & q_0 \end{pmatrix} \mathbf{v} &= \|\mathbf{z}\|_2 \|\mathbf{v}\|_2^2 + 2 \text{Re} \left(v_N \sum_{i=0}^{N-1} z_i v_i^* \right) , \\ &\geq \|\mathbf{z}\|_2 \|\mathbf{v}\|_2^2 - 2|v_N| |\langle \mathbf{z}, \mathbf{v}' \rangle| , \\ &\geq \|\mathbf{z}\|_2 \|\mathbf{v}\|_2^2 - 2|v_N| \|\mathbf{z}\|_2 \|\mathbf{v}'\|_2 , \\ &\geq \|\mathbf{z}\|_2 (\|\mathbf{v}'\|_2^2 - 2|v_N| \|\mathbf{v}'\|_2 + |v_N|^2) , \\ &\geq \|\mathbf{z}\|_2 (\|\mathbf{v}'\|_2 - |v_N|)^2 , \\ &\geq 0 . \end{aligned}$$

Then $\mathbf{q} = (\|\mathbf{z}\|_2, 0, \dots, 0)^\top \in A(\mathbf{z})$ and $q_0 = \|\mathbf{z}\|_2$, which means that $A(\mathbf{z})$ is nonempty and the set $\{\ell(\mathbf{q}) : \mathbf{q} \in A(\mathbf{z})\} \subset \mathbb{R}^+$ is nonempty, so it admits a lower bound less than or equal to $\|\mathbf{z}\|_2$, hence

$$0 \leq \text{SDP}(\mathbf{z}) = \inf_{\mathbf{q} \in A(\mathbf{z})} \ell(\mathbf{q}) \leq \|\mathbf{z}\|_2 .$$

(iii) From (ii), we have

$$\text{SDP}(\mathbf{z}) = \inf_{\mathbf{q} \in A(\mathbf{z})} \ell(\mathbf{q}) = \inf_{B(\mathbf{z})} \ell(\mathbf{q}) ,$$

where

$$B(\mathbf{z}) = \{\mathbf{q} \in A(\mathbf{z}), q_0 \leq \|\mathbf{z}\|_2\} \subset A(\mathbf{z}) .$$

Now, from (89) and (90) we can show that $B(\mathbf{z})$ is bounded since:

$$\forall \mathbf{q} \in A(\mathbf{z}), \quad \|\mathbf{q}\|_2 \leq \sum_{k=1}^r d_k \|\mathbf{a}(f_k)\|_2 \leq \sqrt{N} \sum_{k=1}^r d_k = \sqrt{N} q_0 ,$$

hence

$$\forall \mathbf{q} \in B(\mathbf{z}), \quad \|\mathbf{q}\|_2 \leq \sqrt{N}q_0 \leq \sqrt{N}\|\mathbf{z}\|_2.$$

Consequently,

$$B(\mathbf{z}) \subset B_{\|\cdot\|_2}(0, \|\mathbf{z}\|_2).$$

(iv) Moreover, $A(\mathbf{z}) = \mathbf{T}'_N(\mathbf{z}, \cdot)^{-1}(\mathcal{C})$ is a closed set since the cone of positive matrix \mathcal{C} is closed and the mapping $\mathbf{T}'_N(\mathbf{z}, \cdot)$ is linear, so it is continuous in finite dimension. Thus,

$$B(\mathbf{z}) = A(\mathbf{z}) \cap \{\mathbf{q} \in \mathbb{R}^+ \times \mathbb{C}^{N-1} : q_0 \leq \|\mathbf{z}\|_2\}$$

is a closed set, as the intersection of two closed sets.

(v) From (iii) and (iv), we conclude that $B(\mathbf{z})$ is a compact set and that the function ℓ , which is linear and then continuous, achieves its minimum on $B(\mathbf{z})$ so on $A(\mathbf{z})$, which proves the result (91).

Appendix D. Proof of Proposition 4.

The proof of the direct implication is straightforward. Let us consider the converse one.

By Theorem 2, since $\forall n, \mathbf{T}_M(\mathbf{l}_n) \succcurlyeq 0$ and is of rank one, then there exists $\gamma_n \geq 0$ and $f_n \in [0, 1[$ such that

$$\mathbf{l}_n[m] = \gamma_n \exp(j2\pi f_n m).$$

Since we assume $\forall n, \hat{\mathbf{x}}[0, n] = \hat{\mathbf{x}}[0, 0] = c_1$, then $\mathbf{l}_n[0] = \mathbf{l}_0[0] = c_1$, we have

$$(92) \quad \mathbf{l}_n[m] = c_1 \exp(j2\pi f_n m).$$

Let m be fixed. The Prony matrix $\mathbf{P}_1(\mathbf{t}_m)$ of size $2 \times (N-1)$

$$\mathbf{P}_1(\mathbf{t}_m) = \begin{pmatrix} \mathbf{t}_m[1] & \mathbf{t}_m[0] \\ \vdots & \vdots \\ \mathbf{t}_m[N-1] & \mathbf{t}_m[N-2] \end{pmatrix}$$

is of rank one, consequently there exists $\lambda_m \in \mathbb{C}$ such that

$$\mathbf{t}_m[n+1] = \lambda_m \mathbf{t}_m[n], \quad \forall 0 \leq n \leq N-2.$$

Thus,

$$\mathbf{t}_m[n] = \lambda_m^n \mathbf{t}_m[0], \quad \forall 0 \leq n \leq N-1.$$

From (92) $\mathbf{t}_m[0] = \mathbf{l}_0[m] = c_1 \exp(j2\pi f_0 m)$, $\mathbf{t}_m[1] = \mathbf{l}_1[m] = c_1 \exp(j2\pi f_1 m)$ and then

$$\lambda_m = \frac{\mathbf{t}_m[1]}{\mathbf{t}_m[0]} = \frac{\ell_1[m]}{\ell_0[m]} = \exp(j2\pi(f_1 - f_0)m).$$

Therefore, we have

$$\begin{aligned} \mathbf{t}_m[n] &= \lambda_m^n \mathbf{t}_m[0], \\ &= \exp(j2\pi(f_1 - f_0)m)^n c_1 \exp(j2\pi f_0 m), \\ &= c_1 \exp[j2\pi((f_1 - f_0)n + f_0)m]. \end{aligned}$$

Appendix E. Proof of Proposition 8.

For $\mathbf{z} = (z_0, \dots, z_{N-1}) \in \mathbb{R} \times \mathbb{C}^{N-1}$ and $\mathbf{M} \in \mathcal{T}_N \subset \mathcal{M}_N$ a Hermitian Toeplitz matrix of dimension N . We have:

$$\begin{aligned}
\langle \mathbf{T}_N(\mathbf{z}), \mathbf{M} \rangle_{\mathcal{M}} &= \sum_{0 \leq i, j \leq N-1} [\mathbf{T}_N(\mathbf{z})]_{ij}^* \mathbf{M}_{ij} , \\
&= \sum_{0 \leq i \leq j \leq N-1} z_{j-i} \mathbf{M}_{ij} + \sum_{0 \leq j < i \leq N-1} z_{i-j}^* \mathbf{M}_{ij} , \\
&\stackrel{(*)}{=} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1-k} z_k \mathbf{M}_{l,l+k} + \sum_{k=1}^{N-1} \sum_{l=0}^{N-1-k} z_k^* \mathbf{M}_{l+k,l} , \\
&\stackrel{(**)}{=} z_0 \left(\sum_{l=0}^{N-1} \mathbf{M}_{l,l} \right) + 2\text{Re} \left\{ \sum_{k=1}^{N-1} z_k^* \left(\sum_{l=0}^{N-1-k} \mathbf{M}_{l+k,l} \right) \right\} ,
\end{aligned}$$

with $(*)$ a change of variable $k \leftarrow j - i$ and $(**)$ using that $\mathbf{M}_{l,l+k} = \mathbf{M}_{l+k,l}^*$.

Then, by writing

$$\mathbf{T}'_N(\mathbf{z}, \mathbf{q}) = \left(\begin{array}{c|c} \mathbf{0} & \begin{matrix} z_0 \\ \vdots \\ z_{N-1} \end{matrix} \\ \hline z_0^* & \cdots & z_{N-1}^* & 0 \end{array} \right) + \left(\begin{array}{c|c} \mathbf{T}_N(\mathbf{q}) & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline 0 & \cdots & 0 & q_0 \end{array} \right)$$

we obtain as well for $\mathbf{M} \in \mathcal{T}_{N+1}$:

$$\begin{aligned}
\langle \mathbf{T}'_N(\mathbf{z}, \mathbf{q}), \mathbf{M} \rangle_{\mathcal{M}} &= 2\text{Re} \left\{ \sum_{k=0}^{N-1} z_k^* \mathbf{M}_{N+1,k} \right\} \\
&\quad + q_0 \left(\sum_{l=0}^N \mathbf{M}_{l,l} \right) + 2\text{Re} \left\{ \sum_{k=1}^{N-1} q_k^* \left(\sum_{l=0}^{N-1-k} \mathbf{M}_{l+k,l} \right) \right\} .
\end{aligned}$$

Consequently, the adjoint of the operator

$$\mathbf{T}_{M+1} : (\mathcal{X}_t, \langle \cdot, \cdot \rangle_{\mathcal{X}_t}) \rightarrow (\mathcal{T}_{M+1}, \langle \cdot, \cdot \rangle_{\mathcal{M}}) ,$$

with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{X}_t}$ defined in (35), when applied to a matrix $\mathbf{M}^{(2)} \in \mathcal{T}_{M+1}$, yields the vector

$$\mathbf{z}_2 = \mathbf{T}_{M+1}^* \mathbf{M}^{(2)} \in \mathbb{R} \times \mathbb{C}^M ,$$

whose components are:

$$\mathbf{z}_2[k] = \sum_{l=0}^{M-k} \mathbf{M}_{l+k,l}^{(2)} , \quad \forall k = 0, \dots, M .$$

Similarly, the adjoint of the operator

$$\mathbf{T}'_{H_S} : (\mathcal{X}_t \times \mathcal{Q}_t, \langle \cdot, \cdot \rangle_{\mathcal{X}_t} + \langle \cdot, \cdot \rangle_{\mathcal{Q}_t}) \rightarrow (\mathcal{T}_{H_S+1}, \langle \cdot, \cdot \rangle_{\mathcal{M}}) ,$$

with the inner products $\langle \cdot, \cdot \rangle_{\mathcal{X}_t}$ and $\langle \cdot, \cdot \rangle_{\mathcal{Q}_t}$ defined in (36)-(37), when applied to $\mathbf{M}^{(2)} \in \mathcal{T}_{H_S+1}$, yields the pair of vectors

$$(\mathbf{z}_1, \mathbf{q}_1) = \mathbf{T}_{H_S}^* \mathbf{M}^{(1)} \in \mathbb{C}^{H_S} \times (\mathbb{R} \times \mathbb{C}^{H_S-1}) ,$$

whose components are:

$$\mathbf{z}_1[k] = \mathbf{M}_{H_S+1,k}^{(1)}, \quad \mathbf{q}_1[k] = \sum_{l=0}^{H_S-1-k} \mathbf{M}_{l+k,l}^{(1)} + \delta_k \mathbf{M}_{H_S,H_S}^{(1)}, \quad \forall k = 0, \dots, H_S - 1.$$

Appendix F. Proof of Proposition 9.

First, let us determine the operator norm $\|\mathbf{T}_{M+1}\|^2 = \sup_{\mathbf{z} \in \mathcal{X}_t} \frac{\|\mathbf{T}_{M+1}(\mathbf{z})\|_F^2}{\|\mathbf{z}\|_{\mathcal{X}_t}^2}$.

By definition, we have $\|\mathbf{z}\|_{\mathcal{X}_t}^2 = z_0^2 + 2|z_1|^2 + \dots + 2|z_M|^2$. Moreover, we get:

$$\|\mathbf{T}_{M+1}(\mathbf{z})\|_F^2 = (M+1)z_0^2 + 2M|z_1|^2 + 2(M-1)|z_2|^2 + \dots + 2|z_M|^2 \leq (M+1)\|\mathbf{z}\|_{\mathcal{X}_t}^2,$$

with equality when $\mathbf{z} = (1, 0, \dots, 0)$, hence

$$(93) \quad \|\mathbf{T}_{M+1}\|^2 = M+1.$$

Let us now decompose the operator \mathbf{T}'_{H_S} as follows:

$$\mathbf{T}'_{H_S}(\mathbf{z}, \mathbf{q}) = \underbrace{\left(\begin{array}{c|c} \mathbf{0} & \begin{matrix} z_0 \\ \vdots \\ z_{H_S-1} \end{matrix} \\ \hline \begin{matrix} z_0^* & \dots & z_{H_S-1}^* \end{matrix} & 0 \end{array} \right)}_{\mathbf{T}_1(\mathbf{z})} + \underbrace{\left(\begin{array}{c|c} \mathbf{T}_{H_S}(\mathbf{q}) & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \\ \hline 0 & \dots & 0 & q_0 \end{array} \right)}_{\mathbf{T}_2(\mathbf{q})}.$$

We directly have $\|\mathbf{T}_1(\mathbf{z})\|_F^2 = \|\mathbf{z}\|_{\mathcal{X}_t}^2$, that is $\|\mathbf{T}_1\| = 1$. Besides, we have

$$\|\mathbf{T}_2(\mathbf{q})\|_F^2 = (H_S+1)q_0^2 + 2(H_S-1)|q_1|^2 + \dots + 2|q_{H_S-1}|^2 \leq (H_S+1)\|\mathbf{q}\|_{\mathcal{Q}_t}^2,$$

with equality when $\mathbf{q} = (1, 0, \dots, 0)$, hence $\|\mathbf{T}_2\|^2 = H_S+1$.

Now, we have

$$\begin{aligned} \frac{\|\mathbf{T}'_{H_S}(\mathbf{z}, \mathbf{q})\|_F^2}{\|(\mathbf{z}, \mathbf{q})\|_{\mathcal{X}_t \times \mathcal{Q}_t}^2} &= \frac{\|\mathbf{T}_1(\mathbf{z})\|_F^2 + \|\mathbf{T}_2(\mathbf{q})\|_F^2}{\|\mathbf{z}\|_{\mathcal{X}_t}^2 + \|\mathbf{q}\|_{\mathcal{Q}_t}^2}, \\ &\leq \frac{\|\mathbf{T}_1\|^2 \|\mathbf{z}\|_{\mathcal{X}_t}^2 + \|\mathbf{T}_2\|^2 \|\mathbf{q}\|_{\mathcal{Q}_t}^2}{\|\mathbf{z}\|_{\mathcal{X}_t}^2 + \|\mathbf{q}\|_{\mathcal{Q}_t}^2}, \\ &\leq \alpha \|\mathbf{T}_1\|^2 + (1-\alpha) \|\mathbf{T}_2\|^2, \\ &\leq \max(\|\mathbf{T}_1\|^2, \|\mathbf{T}_2\|^2) = \|\mathbf{T}_2\|^2, \end{aligned}$$

with $\alpha = \frac{\|\mathbf{q}\|_{\mathcal{Q}_t}^2}{\|\mathbf{z}\|_{\mathcal{X}_t}^2 + \|\mathbf{q}\|_{\mathcal{Q}_t}^2}$ and is achieved when $\mathbf{z} = \mathbf{0}$ and $\mathbf{q} = (1, 0, \dots, 0)$, hence

$$\|\mathbf{T}'_{H_S}\|^2 = H_S+1.$$

We are now able to derive the operator norms (48):

$$\begin{aligned} \left\| \mathbf{L}^{(1)}(\mathbf{X}) \right\|_{(1)}^2 &= \sum_{m=1}^M \left\| \mathbf{L}_m^{(1)}(\mathbf{X}) \right\|_{\mathbf{F}}^2, \\ &\leq \left\| \mathbf{T}'_{H_S} \right\|^2 \sum_{m=1}^M \left(\left\| \hat{\mathbf{x}}[m, :] \right\|_{\mathcal{X}_t}^2 + \left\| \mathbf{q}[m, :] \right\|_{\mathcal{Q}_t}^2 \right), \\ &\leq \left\| \mathbf{T}'_{H_S} \right\|^2 \left\| \mathbf{X} \right\|_{\mathcal{H}}^2, \end{aligned}$$

with equality when $\hat{\mathbf{x}} = \mathbf{0}$, $\mathbf{q}[0, :] = \mathbf{0}$ and $\mathbf{q}[m, :] = (1, 0, \dots, 0)$ for all $m \in \llbracket 1, M \rrbracket$.

Similarly, we have

$$\begin{aligned} \left\| \mathbf{L}^{(2)}(\mathbf{X}) \right\|_{(2)}^2 &= \sum_{n_2=0}^{H_S-1} \left\| \mathbf{L}_{n_2}^{(2)}(\mathbf{X}) \right\|_{\mathbf{F}}^2, \\ &\leq \left\| \mathbf{T}_{M+1} \right\|^2 \sum_{n_2=0}^{H_S-1} \left\| \hat{\mathbf{x}}[:, n_2] \right\|_{\mathcal{X}_t}^2, \\ &\leq \left\| \mathbf{T}_{M+1} \right\|^2 \left\| \mathbf{X} \right\|_{\mathcal{H}}^2, \end{aligned}$$

with equality when $\mathbf{q} = \mathbf{0}$ and $\hat{\mathbf{x}}[:, n_2] = (1, 0, \dots, 0)$ for all $n_2 \in \llbracket 0, H_S - 1 \rrbracket$.

We conclude that

$$\begin{aligned} \left\| \mathbf{L}^{(1)} \right\|_{(1)}^2 &= \left\| \mathbf{T}'_{H_S} \right\|^2 = H_S + 1, \\ \left\| \mathbf{L}^{(2)} \right\|_{(2)}^2 &= \left\| \mathbf{T}_{M+1} \right\|^2 = M + 1. \end{aligned}$$

Finally,

$$\begin{aligned} \left\| \mathbf{L}(\mathbf{X}) \right\|_{(1,2)}^2 &= \left\| \mathbf{L}^{(1)}(\mathbf{X}) \right\|_{(1)}^2 + \left\| \mathbf{L}^{(2)}(\mathbf{X}) \right\|_{(2)}^2, \\ &\leq \left(\left\| \mathbf{L}^{(1)} \right\|_{(1)}^2 + \left\| \mathbf{L}^{(2)} \right\|_{(2)}^2 \right) \left\| \mathbf{X} \right\|_{\mathcal{H}}^2. \end{aligned}$$

Appendix G. Proof of Proposition 7. Let F_1 and F_2 be the mappings:

$$\begin{aligned} F_1 : \hat{\mathbf{x}} \in \mathcal{X} &\mapsto \frac{1}{2} \left\| \mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}} \right\|_{\mathcal{Y}}^2 \in \mathbb{R}, \\ F_2 : \mathbf{X} = (\hat{\mathbf{x}}, \mathbf{q}) \in \mathcal{H} &\mapsto \hat{\mathbf{x}} \in \mathcal{X}. \end{aligned}$$

Then, $F : \mathcal{H} \mapsto \mathbb{R}$ writes $F = F_1 \circ F_2$ and its differential at \mathbf{X}_0 is:

$$(\mathrm{d}F)_{\mathbf{X}_0}(\mathbf{X}) = (\mathrm{d}F_1)_{F_2(\mathbf{X}_0)} \circ (\mathrm{d}F_2)_{\mathbf{X}_0}(\mathbf{X}).$$

First, we have

$$\begin{aligned} F_1(\hat{\mathbf{x}} + \mathbf{h}) &= \frac{1}{2} \left\| \mathbf{A}(\hat{\mathbf{x}} + \mathbf{h}) - \hat{\mathbf{y}} \right\|_{\mathcal{Y}}^2, \\ &= \frac{1}{2} \left\| \mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}} \right\|_{\mathcal{Y}}^2 + \frac{1}{2} \langle \mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}, \mathbf{A}\mathbf{h} \rangle_{\mathcal{Y}} + \frac{1}{2} \langle \mathbf{A}\mathbf{h}, \mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}} \rangle_{\mathcal{Y}} + \frac{1}{2} \left\| \mathbf{A}\mathbf{h} \right\|_{\mathcal{Y}}^2, \\ &= F_1(\hat{\mathbf{x}}) + \langle \mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}, \mathbf{A}\mathbf{h} \rangle_{\mathcal{Y}} + o(\|\mathbf{h}\|_{\mathcal{X}}), \\ &= F_1(\hat{\mathbf{x}}) + \langle \mathbf{A}^*(\mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{y}}), \mathbf{h} \rangle_{\mathcal{X}} + o(\|\mathbf{h}\|_{\mathcal{X}}), \end{aligned}$$

that is

$$(\mathrm{d}F_1)_{\widehat{\mathbf{x}}}(\mathbf{h}) = \langle \mathbf{A}^*(\mathbf{A}\widehat{\mathbf{x}} - \widehat{\mathbf{y}}), \mathbf{h} \rangle_{\mathcal{X}}.$$

Moreover, F_2 is linear so $(\mathrm{d}F_2)_{\mathbf{X}_0}(\mathbf{X}) = F_2(\mathbf{X})$, hence

$$(\mathrm{d}F)_{\mathbf{X}_0}(\mathbf{X}) = \langle \mathbf{A}^*(\mathbf{A}\widehat{\mathbf{x}}_0 - \widehat{\mathbf{y}}), \widehat{\mathbf{x}} \rangle_{\mathcal{X}} = \left\langle \begin{pmatrix} \mathbf{A}^*(\mathbf{A}\widehat{\mathbf{x}}_0 - \widehat{\mathbf{y}}) \\ \mathbf{0} \end{pmatrix}, \mathbf{X} \right\rangle_{\mathcal{H}},$$

that is

$$\nabla F(\mathbf{X}_0) = \begin{pmatrix} \mathbf{A}^*(\mathbf{A}\widehat{\mathbf{x}}_0 - \widehat{\mathbf{y}}) \\ \mathbf{0} \end{pmatrix}.$$

Consequently,

$$\|\nabla F(\mathbf{X}) - \nabla F(\mathbf{X}')\|_{\mathcal{H}} \leq \|\mathbf{A}^*\mathbf{A}(\widehat{\mathbf{x}} - \widehat{\mathbf{x}}')\|_{\mathcal{X}} \leq \|\mathbf{A}^*\mathbf{A}\| \|\widehat{\mathbf{x}} - \widehat{\mathbf{x}}'\|_{\mathcal{X}}.$$

We get $\beta = \|\mathbf{A}^*\mathbf{A}\| = \|\mathbf{A}\|^2$ and [Lemma 5](#) concludes the proof.

Appendix H. Proof of [Proposition 10](#). Let us compute the differential function of F :

$$\begin{aligned} & F(\mathbf{X}_1 + \mathbf{h}_1, \mathbf{X}_2 + \mathbf{h}_2) \\ &= \frac{1}{2} \langle \mathbf{A}_1\widehat{\mathbf{x}}_1 + \mathbf{A}_2\widehat{\mathbf{x}}_2 + \mathbf{A}_1\mathbf{h}_1 + \mathbf{A}_2\mathbf{h}_2 - \mathbf{y}, \mathbf{A}_1\widehat{\mathbf{x}}_1 + \mathbf{A}_2\widehat{\mathbf{x}}_2 + \mathbf{A}_1\mathbf{h}_1 + \mathbf{A}_2\mathbf{h}_2 - \mathbf{y} \rangle_{\mathcal{M}}, \\ &= F(\mathbf{X}_1, \mathbf{X}_2) + \frac{1}{2} \langle \mathbf{A}_1\mathbf{h}_1, \mathbf{A}_1\widehat{\mathbf{x}}_1 + \mathbf{A}_2\widehat{\mathbf{x}}_2 - \mathbf{y} \rangle_{\mathcal{M}} + \frac{1}{2} \langle \mathbf{A}_1\mathbf{h}_1, \mathbf{A}_2\mathbf{h}_2 \rangle_{\mathcal{M}} + \frac{1}{2} \langle \mathbf{A}_1\mathbf{h}_1, \mathbf{A}_1\mathbf{h}_1 \rangle_{\mathcal{M}} \\ &\quad + \frac{1}{2} \langle \mathbf{A}_2\mathbf{h}_2, \mathbf{A}_1\widehat{\mathbf{x}}_1 + \mathbf{A}_2\widehat{\mathbf{x}}_2 - \mathbf{y} \rangle_{\mathcal{M}} + \frac{1}{2} \langle \mathbf{A}_2\mathbf{h}_2, \mathbf{A}_1\mathbf{h}_1 \rangle_{\mathcal{M}} + \frac{1}{2} \langle \mathbf{A}_2\mathbf{h}_2, \mathbf{A}_2\mathbf{h}_2 \rangle_{\mathcal{M}}, \end{aligned}$$

with

$$|\langle \mathbf{A}_1\mathbf{h}_1, \mathbf{A}_2\mathbf{h}_2 \rangle_{\mathcal{M}}| \leq \|\mathbf{A}_1\| \|\mathbf{A}_2\| \|\mathbf{h}_1\|_{\mathcal{H}_1} \|\mathbf{h}_2\|_{\mathcal{H}_2} = o(\|(\mathbf{h}_1, \mathbf{h}_2)\|_{\mathcal{H}}),$$

so we deduce that

$$\nabla F(\mathbf{X}_1, \mathbf{X}_2) = \frac{1}{2} \begin{pmatrix} \mathbf{A}_1^*(\mathbf{A}_1\widehat{\mathbf{x}}_1 + \mathbf{A}_2\widehat{\mathbf{x}}_2 - \mathbf{y}) \\ \mathbf{A}_2^*(\mathbf{A}_1\widehat{\mathbf{x}}_1 + \mathbf{A}_2\widehat{\mathbf{x}}_2 - \mathbf{y}) \end{pmatrix}.$$

The adjoint operators are $\tilde{\mathbf{A}}^* \mathbf{z} = (\widehat{\mathbf{G}}_2^* \mathbf{z})^* \bar{\mathbf{h}}'$, $(\mathcal{F}_1^{-1})^* = \frac{1}{W} \mathcal{F}_1$ and $(\mathcal{F}_2^{-1})^* = \frac{1}{H} \mathcal{F}_2$ and $\mathbf{S}_1^*(v_{-M}, \dots, v_0, \dots, v_M) = (v_0, \dots, v_M)$ and $\mathbf{S}_2^*(v_{-P}, \dots, v_0, \dots, v_P) = (v_0, \dots, v_P)$.

Let us determine the Lipschitz constant of the gradient ∇F :

$$\begin{aligned} \|\nabla F(\mathbf{X}_1, \mathbf{X}_2) - \nabla F(\mathbf{X}'_1, \mathbf{X}'_2)\|_{\mathcal{X}}^2 &= \frac{1}{4} \|\mathbf{A}_1^*(\mathbf{A}_1(\widehat{\mathbf{x}}_1 - \widehat{\mathbf{x}}'_1) + \mathbf{A}_2(\widehat{\mathbf{x}}_2 - \widehat{\mathbf{x}}'_2))\|_{\mathcal{X}_1}^2 \\ &\quad + \frac{1}{4} \|\mathbf{A}_2^*(\mathbf{A}_1(\widehat{\mathbf{x}}_1 - \widehat{\mathbf{x}}'_1) + \mathbf{A}_2(\widehat{\mathbf{x}}_2 - \widehat{\mathbf{x}}'_2))\|_{\mathcal{X}_2}^2. \end{aligned}$$

We are looking for a majoration of each term. We deal with the first one C_1 , the second C_2 being obtained in the same way. Using the inequality $(a+b)^2 \leq 2a^2 + 2b^2$:

$$\begin{aligned} C_1 &\leq \frac{1}{4} (\|\mathbf{A}_1^*\mathbf{A}_1\| \|\widehat{\mathbf{x}}_1 - \widehat{\mathbf{x}}'_1\|_{\mathcal{X}_1} + \|\mathbf{A}_1^*\mathbf{A}_2\| \|\widehat{\mathbf{x}}_2 - \widehat{\mathbf{x}}'_2\|_{\mathcal{X}_2})^2, \\ &\leq \frac{1}{2} \|\mathbf{A}_1^*\mathbf{A}_1\|^2 \|\widehat{\mathbf{x}}_1 - \widehat{\mathbf{x}}'_1\|_{\mathcal{X}_1}^2 + \frac{1}{2} \|\mathbf{A}_1^*\mathbf{A}_2\|^2 \|\widehat{\mathbf{x}}_2 - \widehat{\mathbf{x}}'_2\|_{\mathcal{X}_2}^2. \end{aligned}$$

We have $\|\mathbf{A}\| = \|\mathbf{A}^*\| = 1$, $\|\mathbf{S}_i \hat{\mathbf{x}}_1\|_F = \|\hat{\mathbf{x}}_i\|_{\mathcal{X}}$, for $i \in \{1, 2\}$; that is, $\|\mathbf{S}_i\| = 1$ and $\|\mathcal{F}_i^{-1} \mathbf{v}\|_2^2 = \frac{1}{N^2} \|\mathbf{v}\|_2^2$; that is, $\|\mathcal{F}_i^{-1}\| = \frac{1}{N}$. Hence, $\|\mathbf{A}_1\| \leq \frac{1}{W}$, $\|\mathbf{A}_1^*\| \leq 1$, $\|\mathbf{A}_2\| \leq \frac{1}{H}$ and $\|\mathbf{A}_2\| \leq 1$. Consequently, we get

$$C_1 \leq \frac{1}{2W^2} \|\hat{\mathbf{x}}_1 - \hat{\mathbf{x}}'_1\|_{\mathcal{X}_1}^2 + \frac{1}{2H^2} \|\hat{\mathbf{x}}_2 - \hat{\mathbf{x}}'_2\|_{\mathcal{X}_2}^2$$

and exactly the same majoration for C_2 . Thus, we have

$$\|\nabla F(\mathbf{X}_1, \mathbf{X}_2) - \nabla F(\mathbf{X}'_1, \mathbf{X}'_2)\|_{\mathcal{X}}^2 \leq \beta^2 (\|\mathbf{X}_1 - \mathbf{X}'_1\|_{\mathcal{H}_1}^2 + \|\mathbf{X}_2 - \mathbf{X}'_2\|_{\mathcal{H}_2}^2),$$

with

$$\beta = \frac{1}{\min(W, H)}.$$

REFERENCES

- [1] G. AUBERT, L. BLANC-FÉRAUD, AND D. GRAZIANI, *Analysis of a new variational model to restore point-like and curve-like singularities in imaging*, Applied Mathematics & Optimization, 67 (2013), pp. 73–96.
- [2] D. BATENKOV, *Accurate solution of near-colliding Prony systems via decimation and homotopy continuation*, Theoretical Computer Science, 681 (2017), pp. 27–40.
- [3] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex analysis and monotone operator theory in Hilbert spaces*, Springer Science & Business Media, 2011.
- [4] B. N. BHASKAR, G. TANG, AND B. RECHT, *Atomic norm denoising with applications to line spectral estimation*, IEEE Transactions on Information Theory, 61 (2013), pp. 5987–5999.
- [5] T. BLU, *The generalized annihilation property—A tool for solving finite rate of innovation problems*, in Proc. of Int. Workshop on Sampling Theory and Applications (SampTA), May 2009.
- [6] T. BLU, P.-L. DRAGOTTI, M. VETTERLI, P. MARZILIANO, AND L. COULOT, *Sparse sampling of signal innovations*, IEEE Signal Processing Magazine, 25 (2008), pp. 31–40.
- [7] N. BOYD, G. SCHIEBINGER, AND B. RECHT, *The alternating descent conditional gradient method for sparse inverse problems*, SIAM Journal on Optimization, 27 (2017), pp. 616–639.
- [8] S. BOYD AND L. VANDENBERGHE, *Convex optimization*, Cambridge University Press, 2004.
- [9] K. BREDIES, K. KUNISCH, AND T. POCK, *Total generalized variation*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 492–526.
- [10] Y. BRESLER AND A. MACOVSKI, *Exact maximum likelihood parameter estimation of superimposed exponential signals in noise*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 34 (1986), pp. 1081–1089.
- [11] M. BRIANI, A. CUYT, AND W.-S. LEE, *VEXPA: Validated EXponential Analysis through regular sub-sampling*, arXiv preprint arXiv:1709.04281, (2017).
- [12] E.-M. BRINKMANN, M. BURGER, J. RASCH, AND C. SUTOUR, *Bias reduction in variational regularization*, Journal of Mathematical Imaging and Vision, 59 (2017), pp. 534–566.
- [13] E. J. CANDÈS AND C. FERNANDEZ-GRANDA, *Towards a mathematical theory of super-resolution*, Communications on Pure and Applied Mathematics, 67 (2014), pp. 906–956.
- [14] C. CARATHÉODORY, *Über den Variabilitätsbereich der Fourier'schen Konstanten von positiven harmonischen Funktionen*, Rendiconti del Circolo Matematico di Palermo (1884-1940), 32 (1911), pp. 193–217.
- [15] C. CARATHÉODORY AND L. FEJÉR, *Über den Zusammenhang der Extremen von harmonischen Funktionen mit ihren Koeffizienten und über den Picard-Landau'schen Satz*, Rendiconti del Circolo Matematico di Palermo (1884-1940), 32 (1911), pp. 218–239.
- [16] A. CHAMBOLLE, V. CASELLES, D. CREMERS, M. NOVAGA, AND T. POCK, *An introduction to total variation for image analysis*, Theoretical foundations and numerical methods for sparse recovery, 9 (2010), p. 227.
- [17] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, 40 (2011), pp. 120–145.
- [18] Y. CHAN, J. LAVOIE, AND J. PLANT, *A parameter estimation approach to estimation of frequencies of sinusoids*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 29 (1981), pp. 214–219.

- [19] V. CHANDRASEKARAN, B. RECHT, P. A. PARRILO, AND A. S. WILLSKY, *The convex geometry of linear inverse problems*, Foundations of Computational mathematics, 12 (2012), pp. 805–849.
- [20] C. CHEN, P. MARZILIANO, AND A. C. KOT, *2D finite rate of innovation reconstruction method for step edge and polygon signals in the presence of noise*, IEEE Transactions on Signal Processing, 60 (2012), pp. 2851–2859.
- [21] G. CHIERCHIA, N. PUSTELNIK, B. PESQUET-POPESCU, AND J.-C. PESQUET, *A nonlocal structure tensor-based approach for multicomponent image recovery problems*, IEEE Transactions on Image Processing, 23 (2014), pp. 5531–5544.
- [22] P. L. COMBETTES AND J.-C. PESQUET, *Proximal splitting methods in signal processing*, in Fixed-point algorithms for inverse problems in science and engineering, Springer, 2011, pp. 185–212.
- [23] L. CONDAT, *A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, Journal of Optimization Theory and Applications, 158 (2013), pp. 460–479.
- [24] L. CONDAT, *Semi-local total variation for regularization of inverse problems.*, in Proc. of European Signal Processing Conf. (EUSIPCO), 2014, pp. 1806–1810.
- [25] L. CONDAT, *Atomic norm minimization for decomposition into complex exponentials*, research report, GIPSA-Lab, Grenoble, France, (2017).
- [26] L. CONDAT, *Discrete total variation: New definition and minimization*, SIAM Journal on Imaging Sciences, 10 (2017), pp. 1258–1290.
- [27] L. CONDAT, J. BOULANGER, N. PUSTELNIK, S. SAHNOUN, AND L. SENGMANIVONG, *A 2-D spectral analysis method to estimate the modulation parameters in structured illumination microscopy*, in Proc. of IEEE Int. Symp. on Biomedical Imaging (ISBI), 2014, pp. 604–607.
- [28] L. CONDAT AND A. HIRABAYASHI, *Cadizow denoising upgraded: A new projection method for the recovery of Dirac pulses from noisy linear measurements*, Sampling Theory in Signal and Image Processing, 14 (2015), pp. 17–47.
- [29] A. CUYT AND W.-S. LEE, *Multivariate exponential analysis from the minimal number of samples*, Advances in Computational Mathematics, (2017), pp. 1–16.
- [30] E. DAVIES, *Simple two-stage method for the accurate location of Hough transform peaks*, IEE Proceedings E (Computers and Digital Techniques), 139 (1992), pp. 242–248.
- [31] S. R. DEANS, *Hough transform from the Radon transform*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 3 (1981), pp. 185–188.
- [32] C.-A. DELEDALLE, N. PAPADAKIS, AND J. SALMON, *On debiasing restoration algorithms: applications to total-variation and nonlocal-means*, in Proc. of Int. Conf. on Scale Space and Variational Methods in Computer Vision (SSVM), 2015, pp. 129–141.
- [33] C.-A. DELEDALLE, N. PAPADAKIS, J. SALMON, AND S. VAITER, *Clear: Covariant least-square refitting with applications to image restoration*, SIAM Journal on Imaging Sciences, 10 (2017), pp. 243–284.
- [34] B. DIEDERICHS AND A. ISKE, *Parameter estimation for bivariate exponential sums*, in Proc. of Int. Workshop on Sampling Theory and Applications (SampTA), 2015 International Conference on, 2015, pp. 493–497.
- [35] D. L. DONOHO, *Compressed sensing*, IEEE Transactions on information theory, 52 (2006), pp. 1289–1306.
- [36] C. DOSSAL, V. DUVAL, AND C. POON, *Sampling the Fourier transform along radial lines*, SIAM Journal on Numerical Analysis, 55 (2017), pp. 2540–2564.
- [37] P. L. DRAGOTTI, M. VETTERLI, AND T. BLU, *Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang–Fix*, IEEE Transactions on Signal Processing, 55 (2007), pp. 1741–1757.
- [38] A. EFTEKHARI, T. BENDORY, AND G. TANG, *Stable super-resolution of images*, arXiv preprint arXiv:1805.09513, (2018).
- [39] A. EFTEKHARI, J. TANNER, A. THOMPSON, B. TOADER, AND H. TYAGI, *Sparse non-negative super-resolution-simplified and stabilised*, arXiv preprint arXiv:1804.01490, (2018).
- [40] A. EFTEKHARI AND A. THOMPSON, *A bridge between past and present: Exchange and conditional gradient methods are equivalent*, arXiv preprint arXiv:1804.10243, (2018).
- [41] M. ELAD, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer Publishing Company, Incorporated, 1st ed., 2010.
- [42] M. FATEMI, A. AMINI, AND M. VETTERLI, *Sampling and reconstruction of shapes with algebraic boundaries*, IEEE Transactions on Signal Processing, 64 (2016), pp. 5807–5818.
- [43] C. FERNANDEZ-GRANDA, *Super-resolution and compressed sensing*, SIAM News, 46 (2013).
- [44] I. M. GELFAND AND G. E. SHILOV, *Generalized functions, Volume 1: Properties and operations*,

- Fizmatgiz, Moscow, 1958, English transl., Academic Press, New York, (1964).
- [45] P. HOUGH, *Method and means for recognizing complex patterns*, Dec. 18 1962. US Patent 3,069,654.
 - [46] J. ILLINGWORTH AND J. KITTLER, *A survey of the Hough transform*, Computer vision, graphics, and image processing, 44 (1988), pp. 87–116.
 - [47] D. H. JOHNSON, *The application of spectral estimation methods to bearing estimation problems*, Proceedings of the IEEE, 70 (1982), pp. 1018–1028.
 - [48] N. C. K. LASSEN, *Automated Determination of Crystal Orientations from Electron Backscattering Patterns*, PhD thesis, Institute of Mathematical Modelling, 1994.
 - [49] V. F. LEAVERS AND J. F. BOYCE, *The Radon transform and its application to shape parametrization in machine vision*, Image and Vision Computing, 5 (1987), pp. 161–166.
 - [50] S. LEFKIMMATIS, J. P. WARD, AND M. UNSER, *Hessian Schatten-norm regularization for linear inverse problems*, IEEE transactions on image processing, 22 (2013), pp. 1873–1888.
 - [51] M. D. MACLEOD, *Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones*, IEEE Transactions on Signal processing, 46 (1998), pp. 141–148.
 - [52] I. MARAVIC AND M. VETTERLI, *Sampling and reconstruction of signals with finite rate of innovation in the presence of noise*, IEEE Transactions on Signal Processing, 53 (2005), pp. 2788–2805.
 - [53] R. MARCHANT AND P. JACKWAY, *Using super-resolution methods to solve a novel multi-sinusoidal signal model*, in Proc. of Int. Conf. on Digital Image Computing: Techniques and Applications (DICTA), 2013, pp. 1–8.
 - [54] R. MARCHANT AND P. JACKWAY, *A sinusoidal image model derived from the circular harmonic vector*, Journal of Mathematical Imaging and Vision, 57 (2017), pp. 138–163, doi:10.1007/s10851-016-0671-7, <https://doi.org/10.1007/s10851-016-0671-7>.
 - [55] I. MARKOVSKY, *Low rank approximation: algorithms, implementation, applications*, Springer Science & Business Media, 2011.
 - [56] P. MUKHOPADHYAY AND B. B. CHAUDHURI, *A survey of Hough transform*, Pattern Recognition, 48 (2015), pp. 993–1010.
 - [57] L. M. MURPHY, *Linear feature detection and enhancement in noisy images via the Radon transform*, Pattern Recognition Letters, 4 (1986), pp. 279–284.
 - [58] N. NACEREDDINE, S. TABBONE, AND D. ZIOU, *Robustness of Radon transform to white additive noise: general case study*, Electronics Letters, 50 (2014), pp. 1063–1065.
 - [59] G. ONGIE AND M. JACOB, *Off-the-grid recovery of piecewise constant images from few Fourier samples*, SIAM Journal on Imaging Sciences, 9 (2016), pp. 1004–1041.
 - [60] L. ONURAL, *Impulse functions over curves and surfaces and their applications to diffraction*, Journal of mathematical analysis and applications, 322 (2006), pp. 18–27.
 - [61] H. PAN, T. BLU, AND P. L. DRAGOTTI, *Sampling curves with finite rate of innovation*, IEEE Transactions on Signal Processing, 62 (2014), pp. 458–471.
 - [62] N. PARIKH, S. P. BOYD, ET AL., *Proximal algorithms*, Foundations and Trends in optimization, 1 (2014), pp. 127–239.
 - [63] G. PEYRÉ, S. BOUGLEUX, AND L. COHEN, *Non-local regularization of inverse problems*, in Proc. of European Conf. on Computer Vision (ECCV), 2008, pp. 57–68.
 - [64] G. PLONKA AND M. WISCHERHOFF, *How many Fourier samples are needed for real function reconstruction?*, Journal of Applied Mathematics and Computing, 42 (2013), pp. 117–137.
 - [65] K. POLISANO, M. CLAUSEL, L. CONDAT, AND V. PERRIER, *Modélisation de textures anisotropes par la transformée en ondelettes monogènes et super-résolution de lignes 2-D*, PhD thesis, Université Grenoble Alpes, 2017.
 - [66] K. POLISANO, L. CONDAT, M. CLAUSEL, AND V. PERRIER, *Convex super-resolution detection of lines in images*, in Proc. of European Signal Processing Conf. (EUSIPCO), Aug. 2016, pp. 336–340.
 - [67] D. POTTS AND M. TASCHÉ, *Parameter estimation for exponential sums by approximate Prony method*, Signal Processing, 90 (2010), pp. 1631–1642.
 - [68] R. PRONY, *Essai expérimental et analytique : sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l’alkool, à différentes températures*, Journal de l’Ecole Polytechnique Floréal et Plairial, an III, Volume 1, cahier 22 (1795), pp. 24–76.
 - [69] B. G. QUINN AND E. J. HANNAN, *The estimation and tracking of frequency*, vol. 9, Cambridge University Press, 2001.
 - [70] J. RADON, *1.1 über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeitenestimmung*, Classic papers in modern diagnostic radiology, 5 (2005).
 - [71] M. RAHMAN AND K.-B. YU, *Total least squares approach for frequency estimation using linear prediction*, IEEE Transactions on Acoustics, Speech and Signal Processing, 35 (1987), pp. 1440–1454.

- [72] B. RECHT, M. FAZEL, AND P. A. PARRILO, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM review, 52 (2010), pp. 471–501.
- [73] B. RECHT, W. XU, AND B. HASSIBI, *Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization*, in Proc. of IEEE Conf. on Decision and Control (CDC), 2008, pp. 3065–3070.
- [74] A. RENYI, *On projections of probability distributions*, Acta Mathematica Hungarica, 3 (1952), pp. 131–142.
- [75] P. SHUKLA AND P. L. DRAGOTTI, *Sampling schemes for multidimensional signals with finite rate of innovation*, IEEE Transactions on Signal Processing, 55 (2007), pp. 3670–3686.
- [76] H.-C. SO, K. W. CHAN, Y.-T. CHAN, AND K. HO, *Linear prediction approach for efficient frequency estimation of multiple real sinusoids: algorithms and analyses*, IEEE Transactions on Signal Processing, 53 (2005), pp. 2290–2305.
- [77] J.-L. STARCK, F. MURTAGH, AND J. M. FADILI, *Sparse image and signal processing: wavelets, curvelets, morphological diversity*, Cambridge university press, 2010.
- [78] P. STOICA, *List of references on spectral line analysis*, Signal Processing, 31 (1993), pp. 329–340.
- [79] P. STOICA AND R. MOSES, *Spectral Analysis of Signals*, Prentice Hall, NJ, 2005.
- [80] P. STOICA, R. L. MOSES, B. FRIEDLANDER, AND T. SODERSTROM, *Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 37 (1989), pp. 378–392.
- [81] T. STROHMER, *Measure what should be measured: progress and challenges in compressive sensing*, IEEE Signal Processing Letters, 19 (2012), pp. 887–893.
- [82] G. TANG, B. N. BHASKAR, P. SHAH, AND B. RECHT, *Compressed sensing off the grid*, IEEE Transactions on Information Theory, 59 (2013), pp. 7465–7490.
- [83] O. TOEPLITZ, *Zur Theorie der quadratischen und bilinearen Formen von unendlichvielen Veränderlichen*, Mathematische Annalen, 70 (1911), pp. 351–376.
- [84] D. W. TUFTS AND R. KUMARESAN, *Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood*, Proceedings of the IEEE, 70 (1982), pp. 975–989.
- [85] J. A. URIGHEN, Y. C. ELДАР, AND P. DRAGOTTI, *Sampling at the rate of innovation: Theory and applications*, Compressed Sensing: Theory and Applications, (2012), p. 148.
- [86] L. VANDENBERGHE AND S. BOYD, *Semidefinite programming*, SIAM review, 38 (1996), pp. 49–95.
- [87] M. VASCONCELOS, *On the Radon transform of sampled functions*, PhD thesis, Division of Engineering and Applied Sciences, Harvard University, 2003.
- [88] S. YU, J. MA, AND S. OSHER, *Geometric mode decomposition.*, Inverse Problems & Imaging, 12 (2018).